MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0047
Measure Title: Asthma: Pharmacologic Therapy for Persistent Asthma
Measure Steward: The American Academy of Asthma Allergy and Immunology
Brief Description of Measure: Percentage of patients aged 5 years and older with a diagnosis of persistent asthma who were prescribed long-term control medication

Three rates are reported for this measure:
1. Patients prescribed inhaled corticosteroids (ICS) as their long term control medication
2. Patients prescribed other alternative long term control medications (non-ICS)
3. Total patients prescribed long-term control medication

Developer Rationale: This measure promotes the use of long-term control medications for the treatment of persistent asthma. Long-term control medications for the treatment of persistent asthma are the most important type of treatment for most people with asthma. When these medications are taken regularly they can control chronic symptoms and prevent asthma attacks. These medications reduce the underlying inflammation characteristic of asthma, which may result in clinical benefits including reduction in severity of symptoms, improvement in asthma control and quality of life, diminished airway hyper-responsiveness, prevention of exacerbations, and reduction in courses of systemic corticosteroids, ED care, hospitalizations, and deaths due to asthma. Improvement in nationwide prescription of long-term (asthma) control medication for patients with persistent asthma is expected to decrease the number of hospitalizations, reduce healthcare related expenses, and result in improvements as described above.[8]

Inhaled corticosteroids are the preferred controller agent for patients with persistent asthma; however, in patients with mild persistent asthma, anti-leukotriene monotherapy may be an acceptable alternative. [9]

The national asthma guidelines recommend that long-term control medications be taken daily on a long-term basis to achieve and maintain control of persistent asthma.[1] Despite this fact, adherence to daily controller medications remains low and variation in provider practices for asthma medication dosing can be found.[2,3] In a study, nearly half of all asthma patients surveyed indicated that they were not using controller medications and of those, 79% had persistent asthma.[4] Additionally, only 14.3% of those on controllers had well controlled asthma.[4] Uncontrolled disease among the asthma patient population continues to be a US public health concern.[5] The Asthma: Pharmacologic Therapy for Persistent Asthma measure plays an important role in addressing this issue by ensuring that effective clinical care in accordance with national asthma guidelines is delivered. This measure also serves an important role in the Physician Quality Reporting System (PQRS). It is the only asthma measure in the asthma measures group reporting option and it is one of only two asthma measures in the program. According to the CDC, Asthma is a leading chronic illness in the US, affecting an estimated 26 million individuals and costs an estimated $56 billion in medical costs, lost school and works days, and early deaths.[6,7] Asthma measures in national quality reporting programs such as PQRS are needed to ensure quality care is delivered for the asthma patient population.

1.National Heart, Blood and Lung Institute (NHLBI), National Asthma Education and Prevention Program (NAEPP), National Institutes of Health. August 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Full Report 2007. NIH Publication No. 07-4051.
2.Sumino K, Cabana MD. Medication adherence in asthma patients. Curr Opin Pulm Med 2013;19:49–53.
3.Finkelstein JA, Lozano P, Shulruff R, Inui TS, Soumerai SB, Ng M, Weiss KB. Self-reported physician practices for children with asthma: are national guidelines followed? Pediatrics 2000;106: 886–896.
4.Colice GL, Ostrom NK, Geller DE, et al. The CHOICE survey: high rates of persistent and uncontrolled asthma in the United States. Annals of Allergy Asthma & Immunology. 2012 Mar.108(3) 157-U116.

5.Slejko, Julia F., Vahram H. Ghushchyan, Brandon Sucher, Denise R. Globe, Shao-Lee Lin, Gary Globe, and Patrick W. Sullivan. Asthma Control in the United States, 2008-2010: Indicators of Poor Asthma Control. Journal of Allergy and Clinical Immunology 133.6 (2014): 1579-587.
6.National Center for Health Statistics. Summary Health Statistics for US Adults: National Health Interview Survey, 2011. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/nchs/fastats/asthma.htm. Accessed December 8, 2015.
7.National Asthma Control Program. Asthma's Impact on the Nation. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/asthma/impacts_nation/default.htm?s_cid=tw_DrCP274. Accessed December 8, 2015.
8.Diagram 4.1. Relationship of Process Improvements to Outcomes. October 2014. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/asthmaqual/asthmacare/diagram4-1.html
9.Peters SP, Anthonisen N, Castro M, et al. Randomized comparison of strategies for reducing treatment in mild persistent asthma. N Engl J Med. 2007;356(20):2027–2039.

Numerator Statement: Patients who were prescribed long-term control medication
Denominator Statement: All patients aged 5 years and older with a diagnosis of persistent asthma
Denominator Exclusions: Denominator Exceptions:
Documentation of patient reason(s) for not prescribing inhaled corticosteroids or alternative long-term control medication (eg, patient declined, other patient reason)

The AAAAI follows PCPI exception methodology and PCPI distinguishes between measure exceptions and measure exclusions. Exclusions arise when patients who are included in the initial patient or eligible population for a measure do not meet the denominator criteria specific to the intervention required by the numerator. Exclusions are absolute and apply to all patients and therefore are not part of clinical judgment within a measure.

For this measure, exceptions may include patient reason(s) (eg, patient declined). Although this methodology does not require the external reporting of more detailed exception data, the AAAAI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. In further accordance with PCPI exception methodology, the AAAAI advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Registry, Paper Medical Records
Level of Analysis: Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2013

The developer provided the following evidence for this measure:

- The evidence for this measure is based on clinical practice guidelines for the use of long-term medications for patients with persistent asthma from the National Heart, Lung, and Blood Institute (NHLBI), National Asthma Education and Prevention Program (NAEPP), National Institutes of Health. This is a strong recommendation, ranked Category A and includes randomized control trials (RCTs) and expert panels.
- The NHLBI/NAEPP guideline update references a total of 1, 654 articles.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Updates:

- The developer attests there are no new studies that change the evidence—i.e., the new citations listed support the guidelines.
- Several studies also are identified in support of disparities in prevalence of self-reported asthma, trends in medication use, and outcomes.

Exception to evidence: Not applicable

Guidance from the Evidence Algorithm: 1 → 3 → 4→ 5a (highest eligible rating is HIGH)

Question for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

- An estimated 300 million people worldwide suffer from asthma, costing the United States an estimated $56 billion in medical costs, lost school and work days, and early deaths (CDC, 2011).
- According to the CMS Physician Quality Reporting Initiative/System (PQRI/S) 2008 claims data, 46.29% of patients did not meet the measure, which the developer states is evidence of a gap.
- The developer provides information on recent performance scores, but due to changes in the specifications states it is difficult to interpret vis a vis a specific gap.
- Based on its updated testing (CY 2014 data), the developer states the inhaled corticoid steroid rate prescribed for long term control was 88.24 % for all 44 clinics, and the non-inhaled corticosteroid rate long term control medication rate was 71.77% for all sites. The total percentage of patients prescribed long-term control medications for persistent asthma was 99.3%, with some overlap of patients being prescribed BOTH inhaled corticosteroids AND non-inhaled corticosteroids. The developer concludes in the section on updated reliability testing using 2014 data, "There was very small provider to provider or in this Measure clinic to clinic variability."

Disparities

- A 2012 study from CDC's National Center for Health Statistics (NCHA) found the following disparities in prevalence::
  - Disparities existed between adults under and over the age of 65, adults from poor families and those insured by Medicare and/or Medicaid compared to adults privately insured or uninsured.

- Women are more likely to be diagnosed with asthma, hay fever, sinusitis or chronic bronchitis than men

- Also according to the CDC, disparities persist in asthma prevalence and mortality. Disparities exist based on gender, race and ethnicity and income level: African-American adult Medicaid patients with COPD, asthma, or both have a higher mortality and morbidity than their White counterparts. The developer cited several published articles that it states demonstrate that a performance gap persists concerning controller agents for patients with persistent asthma and disparities in their use result in poorer outcomes.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure? Is this measure topped out and so should be considered for Reserve Status?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The evidence basis for the measure has not changed and there is no need for repeat discussion and vote.

**As a maintenance measure, the measure developer has provided a strong case that the underlying evidence for Pharmacologic Therapy for Persistent Asthma has not changed since the last NQF review.

**Strong evidence to support the continued use of the measure.

**This is a process measure supported by multiple clinical studies that examine the link between continued use of long-term inhaled asthma treatments with control of persistent asthma. The process of utilizing a long-term control medication has a direct correlation to increased asthma control.

1b. Performance Gap

Comments:

**There's a gap in care that warrants a national performance measure. A discussion of whether the measure is topped out could be considered.

**The measure developer provided strong evidence of a gap in optimal care as well as variability across providers to justify its continuation as a national performance measure. Though improvements have occurred nationally since first endorsed by the NQF, performance gaps still need to be closed. The developer also provides data that demonstrates disparities based on gender, race & ethnicity, and income.

**Yes. The demonstrated gap is in age, race/ethnicity and in socio-economic status. Lower socio-economic status was shown to have a gap compared to higher socio-economic status.

**Prior claims data shows evidence of a gap; however, specification changes do not necessarily indicate the same. The small sample of clinics queried show a small gap with questionable meaning. Multiple studies have indicated socioeconomic, gender, race, and ethnicity.

1c. High Priority (previously referred to as High Impact)

Comments:

**N/A

**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Registry, Paper Medical Records

Specifications:

- The specifications have changed since the last submission
  - In 2015, the denominator of this measure was modified to take out the age range maximum. The current age range is 5 years old and older. Additionally, the denominator instructions were updated to provide further clarification in accordance with the EPR-3 asthma guidelines.
  - Updates to the numerator of the measure: New tables containing the generic drug names of the applicable medications were developed for each numerator option—inhaled corticosteroids or alternative long-term control medications.
- The numerator of this measure is: *Patients who were prescribed long-term medication.*
- The denominator of this measure is: *All patients age 5 years and older with a diagnosis of persistent asthma.*
- The ICD-9 and ICD-10 codes have been included in the specification details.
- The calculation algorithm is stated in S.18 and appears straightforward.

Question for the Committee:

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Previously, the developer performed data element level validity testing (1 medical center, 86 patients). Per NQF guidance separate reliability testing is not required when validity testing at the data element-level is performed.

Describe any updates to testing

- In the current submission, AAAAI states it updated testing by conducting beta-binomial analysis at the measure-score level.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developer says it conducted beta-binomial analysis on a CY 2014 sample of 1,863 patients from 44 clinic sites through its partner Allergy Partners' EMR system. The developer states the sites are geographically diverse and of diverse practice types.
- The number of patients per site ranged from 1-293 patients, and the number of providers per site from 1- 7.

Results of reliability testing

- The developer reports the following results:
  - Inhaled Corticosteroid Long-Term control medication Reliability = 0.98
  - Non-Inhaled Corticosteroid Long-term Control Medication Reliability = 0.99
  - Combined Long-Term Control Medication Reliability = 0.97
- Although the developer cites the literature review that 0.7 indicates sufficient reliability to distinguish difference among physicians, the developer did not differentiate or do additional analyses based on provider group size (1-6) or #patients/clinic (1-293) in its analyses.

Guidance from the Reliability Algorithm: 1→ 2 → 4 → 5→6 (highest eligible rating is HIGH)

Questions for the Committee:

- *Is the developer's methodology in its updated testing appropriate?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- For previous empirical testing, data from an EHR performance report were compared to data elements found in the medical record and scores calculated manually on visual inspection by trained abstractors.
  - Data sample was 86 patient encounters, from 1 academic medical center located in an urban area.
  - Data analysis included: 1) Percent agreement, and 2) Kappa statistic
    - Reliability: N, % Agreement, Kappa
    - Numerator: 86, 90.1%, 0.00* (-0.6579-0.6579 CI)
    - Denominator: 86, 94.2%, 0.00* (-0.8507-0.8507 CI)
    - The developer explains the Kappa of 0 as follows: *This is an example of the limitation of the Kappa statistic. While the agreement can be 90% or greater, if one classification category dominates, Kappa can be significantly reduced."
- Face validity of the measure score was assessed by a 14-member expert panel, the of a Joint Task Force on Quality and Performance Measures.
  - Expert panel rating of the validity statement: N = 8; Mean rating = 4.875 and 100% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality
- The previous Committee noted concerns about patients receiving combination medications that include ICS not being counted in rate 1 – preferred therapy.

Describe any updates to validity testing

- The developer conducted new face validity testing at the performance score level.
- The reports on a denominator certification and data quality check process for the testing dataset as data element level validity, but provides only a description and no statistical analyses or other details as required by NQF. Accordingly, the validity testing is represented by 1) the previous data element-level validity (above section), and 2) new face validity testing (below section).

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**
☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer conducted new face validity testing at the performance score measure using a group of 29 asthma, allergy, and immunology experts not involved with the measure development.

Validity testing results:

- The mean rating for the new face validity performance score-level assessment was 4.79 out of 5.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

From the developer's initial testing:

- Exceptions included patient reason. Specifications allow for documented patient exception.
- There were no exceptions documented in this project. All sampled patients were able to be assessed.

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Questions for the Committee:

- *None*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer's meaningful difference analysis derived from the initial testing (PQRS 2008)

- Meaningful difference in performance is calculated using an inter-quartile range (IQR) for each indicator.
  Scores on this measure: N = 337; Mean = 53.71%
  10th percentile: 0.0%
  25th percentile: 0.0%
  50th percentile: 100.00%
  75th percentile: 100.00%
  90th percentile: 100.00%
- Elsewhere in the section on updated reliability testing using 2014 data, the developer states, "There was very small provider to provider or in this Measure clinic to clinic variability."

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

- Measure calculated using data collected using two different methods of collection

2b7. Missing Data

- The developer reports this is not applicable.

Guidance from the Validity Algorithm: 1 → 2 → 3 → 4→ 5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The specifications are consistent with the evidence.
**Specifications are consistent.
**Consistency achieved.
**Developer notes that specifications are consistent with evidence in 1A

2a2. Reliability Testing

Comments:

**The results demonstrate sufficient validity so that conclusions about quality can be made and the measure score is an indicator of quality.
**Validity testing show rates of 90.1% for the numerator and 94.2% for the denominator and are sufficient for use in generalizing for

widespread implementation.
**Validity was described in the measure specifications and they are reasonable.
**Sample size of 86 patients from single academic medical center; scores calculated on visual inspection. Validity testing level was data element with face validity only. Algorithm indicates moderate level of validity.

2b2. Validity Testing
Comments:
**n/a
**Exclusions (patient reason) are appropriate. Risk Adjustment n/a. Meaningful differences clearly identifies a quality gap. Missing data N/A.
**The exclusion of refused will require chart audit to collect and confirm unless electronic health records capture and are able to be used for reporting purposes.
**N/A

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**AAAAI states it updated testing by conducting beta-binomial analysis at the measure-score level. Although the developer cites the literature review that 0.7 indicates sufficient reliability to distinguish difference among physicians, the developer did not differentiate or do additional analyses based on provider group size (1-6) or #patients/clinic (1-293) in its analyses. Despite this, methodology appears appropriate and results demonstrate sufficient reliability.
**The data provided demonstrated sufficient reliability to identify performance differences.
**Reliability testing was performed. A measure already in place with updates to coding, etc. Similar to other asthma measures in existence.
**Testing was performed on 1863 patients in 44 clinic sites; reliability levels were above minimum, but developer did not provide any data for differentiation between provide and patient group sizes among clinics evaluated.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic health records (EHRs).

Question for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**All data elements are in defined fields in electronic health records (EHRs).
**All data element are routinely generated in EHRs during the course of care delivery.
**Not all E.H.R.'s may capture the refusal in any easily reportable field. All other data lend itself to electronic reporting.
**All of the data elements can be collected via EHR.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
Public Reporting
- Physician Quality Reporting System

Payment Program
- Physician Quality Reporting System
- The AAAAI Quality Clinical Data Registry in Collaboration with CECity
- The AAAAI Quality Clinical Data Registry in Collaboration with CECity

Professional Certification or Recognition Program
- Asthma Specialist Tool to Help Manage Asthma and Improve Quality (ASTHMA IQ)
- The American Board of Internal Medicine (ABIM) Self-Directed Practice Improvement Module (PIM)

Quality Improvement with Benchmarking
- The American Academy of Allergy Asthma and Immunology Quality Clinical Data Registry in Collaboration with CECity

Improvement results
- Physician Quality Reporting System: According to the 2013 PQRS experience report, in 2013, 641,654 eligible professionals (51%) participated in PQRS
- ASTHMA IQ: Data as of 11/13/2012 – Users (specialist and primary care): 1,133; patients (specialist and primary care): 10,387; and clinics (specialist and primary care): 1,052
- The American Academy of Allergy Asthma and Immunology Quality Clinical Data Registry in Collaboration with CECity (AAAAI QCDR): data on this measure are not available from 2014.

Unexpected findings (positive or negative) during implementation [unexpected findings]
- This measure was one of the top 14 measures where more than 20% of reported instances were for a patient who did not meet the necessary age range. The age range was modified in 2014 to 5 years and older to correct the issue.
- The 2013 PQRS Experience Report indicates in 2011 this measure had an average performance score of 69.1%. In 2013, the average performance score for this measure was 89.4%. The developer speculates the use of this measure in PQRS has improved the rates.

Potential harms: The developer is not aware of any unintended consequences.

Feedback: No feedback provided on QPS. Measure reviewed by MAP for Physician Quality Reporting System (PQRS) and Medicare and Medicaid EHR Incentive Program for Eligible Professionals in 2012 and 2013. MAP voted to retain the measure in PQRS. The measure was also reviewed in Physician Compare in 2014 and Value-Based Payment Modifier (VBPM) Program in 2012 and 2014. In 2014, MAP voted to not support the measure because it does not adequately address any current needs of the Physician Compare and VBPM) program.

***Questions for the Committee**:*
- *During the last Committee review it was noted this is a retooled eMeasure from the meaningful use program. Does the Committee wish to ask developers about use of and results from the eMeasure version?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments:

**Publicly reported measure. Will defer to committee on whether to discuss results and use of eMeasure version.

**This is maintenance measures that is already in use the for Public Reporting (PQRS) and Payment programs (PQRS and the AAAAI Quality Clinical Data Registry) . Performance between 2011 and 2013 has steadily improved. Continued public reporting and payment programs should drive higher quality care, which will also result in lower per capita costs associated with Asthma care. Do not foresee unintended consequences.

**No concerns.

**The measure is currently publicly reported and utilized in accountability programs: Physician quality reporting System; AAAAI Quality clinical Data Registry in Collaboration with CECity; Asthma Specialist Tool to Help manage Asthma and Improve quality; American Board of Internal Medicine Self-Directed Practice Improvement Module

Criterion 5: Related and Competing Measures

Related or competing measures

- 1799: Medication Management for People with Asthma
- 1800: Asthma Medication Ratio

Harmonization

- Measures 1799 and 1800 have not been harmonized.
- The levels of analysis differ (1799 and 1800 are health-plan; 0047 is clinician (individual or group)
- The developer presents the following denominator differences:
  - The three measures are similar in regards to the denominator population of patients with persistent asthma. The denominators differ with respect to the method patients are identified.
  - For measures 1799 and 1800, persistent asthma is defined from administrative data. For measure 0047, persistent asthma is defined based on clinical information.
  - The denominator for measure 0047 includes asthma patients aged 65 and older, a population not reached by measures 1799 and 1800
- The developer notes the following numerator differences:
  - The numerator for 0047 differs from 1799 because inhaled corticosteroids and alternative controllers are reported separately as well as together.
  - The alternative long-term controllers included in the numerator for measure 0047 have not been harmonized with measure 1799.

During the last review the Committee only requested harmonization of measures 0047 and 0036. However, endorsement has been removed from measure 0036.

Pre-meeting public and member comments

- None

NQF #: 0047 NQF Project: Pulmonary Project

1. IMPACT, OPPORTUITY, 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)

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)*:

The focus of this measure is the prescription of long-term control medications for patients with persistent asthma (process). Use of these medications is associated with improved health outcomes, eg reduction in severity of symptoms; improvement in asthma control and quality of life; prevention of exacerbations; reduction in ED care, hospitalizations, and deaths due to asthma.

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

Clinical Practice Guideline

Randomized Controlled Trials

Longitudinal Cohort Studies

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)*:

The referenced guideline recommendations and underlying body of evidence pertain to the effectiveness of long-term control medications for achieving and maintaining control of persistent asthma, and are thus directly relevant to this measure.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: The NHLBI/NAEPP guideline update references a total of 1,654 articles selected for use in updating the guideline. Evidence tables quantifying the studies reviewed for developing the guideline recommendations related to pharmacologic therapy may be found on the NHLBI web site:

http://www.nhlbi.nih.gov/guidelines/asthma/evid_tbls.htm

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):* As indicated by the "Category A" ranking of the evidence for this measure, the body of evidence consists of a substantial number of well-designed RCTs and a rich body of data that provide a consistent pattern of findings in the population for which the recommendation is made. The "A" evidence ranking thus also suggests a high degree of confidence in the benefits of the

recommendation to patients.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Per the evidence grading scale for the NHLBI/NAEPP guideline, the "A" evidence ranking signifies a consistent pattern of findings across the studies used to formulate the recommendation; a "B" ranking would have been assigned had the findings been "somewhat inconsistent." The NHLBI/NAEPP guideline does not provide any more explicit information related to the consistency of the studies underlying the guideline recommendations.

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

Again, the "A" evidence ranking and strong recommendation in the NHLBI/NAEPP guideline suggests a high degree of confidence in the benefits of long-term control medications for patients with persistent asthma. The NHLBI/NAEPP guideline does not otherwise provide any explicit information related to the benefit/harm ratio for this recommendation.

The national asthma guidelines recommend that long-term control medications be taken daily on a long-term basis to achieve and maintain control of persistent asthma.[1] Despite this fact, adherence to daily controller medications remains low and variation in provider practices for asthma medication dosing can be found.[2,3] In a study, nearly half of all asthma patients surveyed indicated that they were not using controller medications and of those, 79% had persistent asthma.[4] Additionally, only 14.3% of those on controllers had well controlled asthma. [4] Uncontrolled disease among the asthma patient population continues to be a US public health concern. [5] The Asthma: Pharmacologic Therapy for Persistent Asthma measure plays an important role in addressing this issue by ensuring that effective clinical care in accordance with national asthma guidelines is delivered. This measure also serves an important role in the Physician Quality Reporting System (PQRS). It is the only asthma measure in the asthma measures group reporting option and it is one of only two asthma measures in the program. According to the CDC, Asthma is a leading chronic illness in the US, affecting an estimated 26 million individuals and costs an estimated $56 billion in medical costs, lost school and works days, and early deaths. [6,7] Asthma measures in national quality reporting programs such as PQRS are needed to ensure quality care is delivered for the asthma patient population.

This measure promotes the use of long-term control medications for the treatment of persistent asthma. Long-term control medications for the treatment of persistent asthma are the most important type of treatment for most people with asthma. When these medications are taken regularly they can control chronic symptoms and prevent asthma attacks. These medications reduce the underlying inflammation characteristic of asthma, which may result in clinical benefits including reduction in severity of symptoms, improvement in asthma control and quality of life, diminished airway hyper-responsiveness, prevention of exacerbations, and reduction in courses of systemic corticosteroids, ED care, hospitalizations, and deaths due to asthma. Improvement in nationwide prescription of long-term (asthma) control medication for patients with persistent asthma is expected to decrease the number of hospitalizations, reduce healthcare related expenses, and result in improvements as described above. [8]

Inhaled corticosteroids are the preferred controller agent for patients with persistent asthma; however, in patients with mild persistent asthma, anti-leukotriene monotherapy may be an acceptable alternative. [9]

Additional cohort/longitudinal studies:

The CHOICE (Comprehensive Survey of Healthcare Professionals and Asthma Patients Offering Insight on Current Treatment Gaps and Emerging Device Options) survey conducted in 2011 found that almost half (490) of the patients participating in the survey were not using controller medications. Of this population, 70% had persistent asthma and 47% had mild or moderate persistent asthma. The survey was conducted over a 12 month period during which time patients not on asthma controllers reported overnight hospitalizations in 7.1% of cases, 12.9% had an ED or outpatient visit and 17.1% had an unscheduled office visit for asthma. Use of an acute care facility was greatest among patients with a higher severity level of asthma and least utilized in patients with well-controlled asthma. The study concluded that many patients with untreated asthma would benefit from long-term controllers and

further noted that better management was needed of patients on long-term controller medications. (10)

A 2013 study of Medical Expenditure Pane(11) Survey data from 2008-2010 examined the national prevalence of self-reported asthma and trends in medication use. Findings included:

- 14.6% of respondents indicated use of 3 or more canisters of quick-relief inhalers in the past 3 months
- Of those using short-acting inhalers, 60% used daily long-term control medication but still depended on the use of short-acting inhalers
- 28% had never used a long-term control medication and 54% of respondents who had experienced a recent asthma exacerbation indicated they have never used long-term control medication

A study published in 2006 found that from 1998-2000 (12):

•43% of patients who reported using a beta2-agonist inhaler more than three times a day on a daily basis were prescribed a longer acting bronchodilator and/or anti-inflammatory agent.

•60% of patients who required chronic treatment with systemic corticosteroids during any 12 month period were prescribed inhaled corticosteroids during that same time period.

Another study evaluated quality of care for children in the United States. From 1998-2000 (13):

•43% of patients who reported using a beta2-agonist inhaler more than three times a day on a daily basis were prescribed a longer acting bronchodilator and/or anti-inflammatory agent.

•8% of patients requiring chronic treatment with oral corticosteroids had a trial of inhaled corticosteroids first

•80% of patients who required frequent bursts of prednisone who were not already on inhaled corticosteroids or cromolyn were started on them

References

1. National Heart, Blood and Lung Institute (NHLBI), National Asthma Education and Prevention Program (NAEPP), National Institutes of Health. August 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Full Report 2007. NIH Publication No. 07-4051.
2. Sumino K, Cabana MD. Medication adherence in asthma patients. Curr Opin Pulm Med 2013;19:49–53.
3. Finkelstein JA, Lozano P, Shulruff R, Inui TS, Soumerai SB, Ng M, Weiss KB. Self-reported physician practices for children with asthma: are national guidelines followed? Pediatrics 2000;106: 886–896.
4. Colice GL, Ostrom NK, Geller DE, et al. The CHOICE survey: high rates of persistent and uncontrolled asthma in the United States. Annals of Allergy Asthma & Immunology. 2012 Mar.108(3) 157-U116.
5. Slejko, Julia F., Vahram H. Ghushchyan, Brandon Sucher, Denise R. Globe, Shao-Lee Lin, Gary Globe, and Patrick W. Sullivan. Asthma Control in the United States, 2008-2010: Indicators of Poor Asthma Control. Journal of Allergy and Clinical Immunology 133.6 (2014): 1579-587.
6. National Center for Health Statistics. *Summary Health Statistics for US Adults: National Health Interview Survey, 2011*. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/nchs/fastats/asthma.htm. Accessed December 8, 2015.
7. National Asthma Control Program. *Asthma's Impact on the Nation*. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/asthma/impacts_nation/default.htm?s_cid=tw_DrCP274. Accessed December 8, 2015.
8. Diagram 4.1. Relationship of Process Improvements to Outcomes. October 2014. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/asthmaqual/asthmacare/diagram4-1.html
9. Peters SP, Anthonisen N, Castro M, et al. Randomized comparison of strategies for reducing treatment in mild persistent asthma. N Engl J Med. 2007;356(20):2027–2039.
10. Colice GL, Ostrom NK, Geller DE, et al. The CHOICE survey: high rates of persistent and uncontrolled asthma in the United States. Annals of Allergy Asthma & Immunology. 2012 Mar.108(3) 157-U116.
11. Slejko, Julia F., Vahram H. Ghushchyan, Brandon Sucher, Denise R. Globe, Shao-Lee Lin, Gary Globe, and Patrick W. Sullivan. "Asthma Control in the United States, 2008-2010: Indicators of Poor Asthma Control." Journal of Allergy and Clinical Immunology 133.6 (2014): 1579-587

12. Technical Appendix to McGlynn EA, A. S., Adams JL, et al. (2006). "Who is at greatest risk for receiving poor quality health care?" N Engl J Med 354: 1147-1156.
13. Mangione-Smith R, D. A., Setodji CM, et al. (2007). "The quality of ambulatory care delivered to children in the United States: Supplementary Appendix." N Engl J Med 357: 1515-1523.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: THIRD EXPERT PANEL ON THE DIAGNOSIS AND MANAGEMENT OF ASTHMA (NHLBI/NAEPP)

William W. Busse, M.D., Chair
University of Wisconsin Medical School
Madison, Wisconsin

Homer A. Boushey, M.D.
University of California–San Francisco
San Francisco, California

Carlos A. Camargo, Jr., M.D., Dr.P.H.
Massachusetts General Hospital
Boston, Massachusetts

David Evans, Ph.D., A.E.-C,
Columbia University
New York, New York

Michael B. Foggs, M.D.
Advocate Health Centers
Chicago, Illinois

Susan L. Janson, D.N.Sc., R.N., A.N.P.,
F.A.A.N.
University of California–San Francisco
San Francisco, California

H. William Kelly, Pharm.D.

University of New Mexico Health Sciences
Center
Albuquerque, New Mexico

Robert F. Lemanske, M.D.
University of Wisconsin Hospital and Clinics
Madison, Wisconsin

Fernando D. Martinez, M.D.
University of Arizona Medical Center
Tucson, Arizona

Robert J. Meyer, M.D.
U.S. Food and Drug Administration
Silver Spring, Maryland

Harold S. Nelson, M.D.
National Jewish Medical and Research
Center
Denver, Colorado

Thomas A. E. Platts-Mills, M.D., Ph.D.
University of Virginia School of Medicine
Charlottesville, Virginia

Michael Schatz, M.D., M.S.
Kaiser-Permanente–San Diego
San Diego, California

Gail Shapiro, M.D.†
University of Washington
Seattle, Washington

Stuart Stoloff, M.D.
University of Nevada School of Medicine

Carson City, Nevada

Stanley J. Szefler, M.D.

National Jewish Medical and Research

Center

Denver, Colorado

Scott T. Weiss, M.D., M.S.

Brigham and Women's Hospital

Boston, Massachusetts

Barbara P. Yawn, M.D., M.Sc.

Olmstead Medical Center

Rochester, Minnesota

†Deceased

Development of the resource document and the guidelines report was funded by the NHLBI, NIH. Expert Panel members completed financial disclosure forms, and the Expert Panel members disclosed relevant financial interests to each other prior to their discussions. Expert Panel members participated as volunteers and were compensated only for travel expenses related to the Expert Panel meetings. Financial disclosure information covering the 3-year period during which the guidelines were developed is provided for each Panel member below.

Dr. Busse has served on the Speakers' Bureaus of GlaxoSmithKline, Merck, Novartis, and Pfizer; and on the Advisory Boards of Altana, Centocor, Dynavax, Genentech/Novartis, GlaxoSmithKline, Isis, Merck, Pfizer, Schering, and Wyeth. He has received funding/grant support for research projects from Astellas, AstraZeneca, Centocor, Dynavax, GlaxoSmithKline, Novartis, and Wyeth. Dr. Busse also has research support from the NIH.

Dr. Boushey has served as a consultant for Altana, Protein Design Lab, and Sumitomo. He has received honoraria from (Boehringer-Ingelheim, Genentech, Merck, Novartis, and Sanofi-Aventis, and funding/grant support for research projects from the NIH.

Dr. Camargo has served on the Speakers' Bureaus of AstraZeneca, GlaxoSmithKline, Merck, and Schering-Plough; and as a consultant for AstraZeneca, Critical Therapeutics, Dey Laboratories, GlaxoSmithKline, MedImmune, Merck, Norvartis, Praxair, Respironics, Schering-Plough, Sepracor, and TEVA. He has received funding/grant support for research projects from a variety of Government agencies and not-for-profit foundations, as well as

AstraZeneca, Dey Laboratories, GlaxoSmithKline, MedImmune, Merck, Novartis, and

Respironics.

Dr. Evans has received funding/grant support for research projects from the NHLBI. Dr. Foggs has served on the Speakers' Bureaus of GlaxoSmithKline, Merck, Pfizer, Sepracor, and UCB Pharma; on the Advisory Boards of Alcon, Altana, AstraZeneca, Critical Therapeutics, Genentech, GlaxoSmithKline, and IVAX; and as consultant for Merck and Sepracor. He has received funding/grant support for research projects from GlaxoSmithKline.

Dr. Janson has served on the Advisory Board of Altana, and as a consultant for Merck. She has received funding/grant support for research projects from the NHLBI.

Dr. Kelly has served on the Speakers' Bureaus of AstraZeneca and GlaxoSmithKline; and on the Advisory Boards of AstraZeneca, MAP Pharmaceuticals, Merck, Novartis, and Sepracor.

Dr. Lemanske has served on the Speakers' Bureaus of GlaxoSmithKline and Merck, and as a consultant for AstraZeneca, Aventis, GlaxoSmithKline, Merck, and Novartis. He has received honoraria from Altana, and funding/grant support for research projects from the NHLBI and NIAID.

Dr. Martinez has served on the Advisory Board of Merck and as a consultant for Genentech, GlaxaSmithKline, and Pfizer. He has received honoraria from Merck.

Dr. Meyer has no relevant financial interests.

Dr. Nelson has served on the Speakers' Bureaus of AstraZeneca, GlaxoSmithKline, Pfizer, and Schering-Plough; and as a consultant for Abbott Laboratories, Air Pharma, Altana Pharma US, Astellas, AstraZeneca, Curalogic, Dey Laboratories, Dynavax Technologies, Genentech/Novartis, GlaxoSmithKline, Inflazyme Pharmaceuticals, MediciNova, Protein Design Laboratories, Sanofi-Aventis, Schering-Plough, and Wyeth Pharmaceuticals. He has received funding/grant support for research projects from Altana, Astellas, AstraZeneca, Behringer, Critical Therapeutics, Dey Laboratories, Epigenesis, Genentech, GlaxoSmithKline, Hoffman LaRoche, IVAX, Medicinova, Novartis, Sanofi-Aventis, Schering-Plough, Sepracor, TEVA, and Wyeth.

Dr. Platts-Mills has served on the Advisory Committee of Indoor Biotechnologies. He has received funding/grant support for a research project from Pharmacia Diagnostics.

Dr. Schatz has served on the Speakers' Bureaus of AstraZeneca, Genentech, GlaxoSmithKline, and Merck; and as a consultant for GlaxoSmithKline on an unbranded asthma initiative. He has received honoraria from AstraZeneca, Genentech, GlaxoSmithKline and Merck. He has received funding/grant support for research projects from GlaxoSmithKline and Merck and Sanofi-Adventis.

Dr. Shapiro† served on the Speakers' Bureaus of AstraZeneca, Genentech, GlaxoSmithKline, IVAX Laboratories, Key Pharmaceuticals, Merck, Pfizer Pharmaceuticals, Schering Corporation, UCB Pharma, and 3M; and as a consultant for Altana, AstraZeneca, Dey Laboratories, Genentech/Novartis, GlaxoSmithKline, ICOS, IVAX Laboratories, Merck, Sanofi-Aventis, and Sepracor. She received funding/grant support for research projects from Abbott, AstraZeneca,

Boehringer Ingelheim, Bristol-Myers-Squibb, Dey Laboratories, Fujisawa Pharmaceuticals, Genentech, GlaxoSmithKline, Immunex, Key, Lederle, Lilly Research, MedPointe Pharmaceuticals, Medtronic Emergency Response Systems, Merck,

Novartis, Pfizer, Pharmaxis, Purdue Frederick, Sanofi-Aventis, Schering, Sepracor, 3M Pharmaceuticals, UCB Pharma, and Upjohn Laboratories.

Dr. Stoloff has served on the Speakers' Bureaus of Alcon, Altana, AstraZeneca, Genentech, GlaxoSmithKline, Novartis, Pfizer, Sanofi-Aventis, and Schering; and as a consultant for Alcon, Altana, AstraZeneca, Dey, Genentech, GlaxoSmithKline, Merck, Novartis, Pfizer, Sanofi-Aventis, and Schering.

Dr. Szefler has served on the Advisory Boards of Altana, AstraZeneca, Genentech,

GlaxoSmithKline, Merck, Novartis, and Sanofi-Aventis; and as a consultant for Altana, AstraZeneca, Genentech, GlaxoSmithKline, Merck, Novartis, and Sanofi-Aventis. He has received funding/grant support for a research project from Ross.

Dr. Weiss has served on the Advisory Board of Genentech, and as a consultant for Genentech and GlaxoSmithKline. He has received funding/grant support for research projects from GlaxoSmithKline.

Dr. Yawn has served on the Advisory Boards of Altana, AstraZeneca, Merck, Sanofi-Aventis, and Schering-Plough. She has received honoraria from Pfizer and Schering-Plough, and funding/grant support for research projects from the Agency for Healthcare Research and Quality, the CDC, the NHLBI, Merck, and Schering-Plough.

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

1c.12 If other, identify and describe the grading scale with definitions: Evidence Category A: Randomized controlled trials (RCTs), rich body of data.

Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.

Evidence Category B: RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.

Evidence Category C: Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

Evidence Category D: Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories.

The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

1c.13 Grade Assigned to the Body of Evidence: Category A

1c.14 Summary of Controversy/Contradictory Evidence: The guideline identifies no contradictory evidence related to this recommendation.

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

None

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

The Expert Panel recommends that long-term control medications be taken daily on a long-term basis to achieve and maintain control of persistent asthma. The most effective long-term-control medications are those that attenuate the underlying inflammation characteristic of asthma. (NHLBI/NAEPP, pg. 216)

The Expert Panel concludes that ICSs are the most potent and consistently effective long-term control medication for asthma. (NHLBI/NAEPP, pg. 216)

The Expert Panel concludes that ICSs are the most effective long-term therapy available for patients who have persistent asthma and, in general, ICSs are well tolerated and safe at the recommended dosages. (NHLBI/NAEPP, pg. 220)

1c.17 Clinical Practice Guideline Citation: National Heart, Blood and Lung Institute (NHLBI), National Asthma Education and Prevention Program (NAEPP), National Institutes of Health. August 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Full Report 2007. NIH Publication No. 07-4051.

1c.18 National Guideline Clearinghouse or other URL: http://www.guideline.gov/browse/by-organization.aspx?orgid=400

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: (see 1c.10)

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

1c.22 If other, identify and describe the grading scale with definitions: In addition to specifying the level of evidence supporting a recommendation, the Expert Panel agreed to indicate the strength of the recommendation. When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel. When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong.

(NHLBI/NAEPP, pg. 7)

1c.23 Grade Assigned to the Recommendation: Strong

1c.24 Rationale for Using this Guideline Over Others: It is the PCPI policy to use guidelines, which are evidence-based, applicable to physicians and other health-care providers, and developed by a national specialty organization or government agency. In addition, the PCPI has now expanded what is acceptable as the evidence base for measures to include documented quality improvement (QI) initiatives or implementation projects that have demonstrated improvement in quality of care.

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: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0047_Evidence_MSF5.0_Data.doc,0047_Evidence_Attachment.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure promotes the use of long-term control medications for the treatment of persistent asthma. Long-term control medications for the treatment of persistent asthma are the most important type of treatment for most people with asthma. When these medications are taken regularly they can control chronic symptoms and prevent asthma attacks. These medications reduce the underlying inflammation characteristic of asthma, which may result in clinical benefits including reduction in severity of symptoms, improvement in asthma control and quality of life, diminished airway hyper-responsiveness, prevention of exacerbations, and reduction in courses of systemic corticosteroids, ED care, hospitalizations, and deaths due to asthma. Improvement in nationwide prescription of long-term (asthma) control medication for patients with persistent asthma is expected to decrease the number of hospitalizations, reduce healthcare related expenses, and result in improvements as described above.[8]

Inhaled corticosteroids are the preferred controller agent for patients with persistent asthma; however, in patients with mild persistent asthma, anti-leukotriene monotherapy may be an acceptable alternative. [9]

The national asthma guidelines recommend that long-term control medications be taken daily on a long-term basis to achieve and maintain control of persistent asthma.[1] Despite this fact, adherence to daily controller medications remains low and variation in provider practices for asthma medication dosing can be found.[2,3] In a study, nearly half of all asthma patients surveyed indicated

that they were not using controller medications and of those, 79% had persistent asthma.[4] Additionally, only 14.3% of those on controllers had well controlled asthma.[4] Uncontrolled disease among the asthma patient population continues to be a US public health concern.[5] The Asthma: Pharmacologic Therapy for Persistent Asthma measure plays an important role in addressing this issue by ensuring that effective clinical care in accordance with national asthma guidelines is delivered. This measure also serves an important role in the Physician Quality Reporting System (PQRS). It is the only asthma measure in the asthma measures group reporting option and it is one of only two asthma measures in the program. According to the CDC, Asthma is a leading chronic illness in the US, affecting an estimated 26 million individuals and costs an estimated $56 billion in medical costs, lost school and works days, and early deaths.[6,7] Asthma measures in national quality reporting programs such as PQRS are needed to ensure quality care is delivered for the asthma patient population.

1.National Heart, Blood and Lung Institute (NHLBI), National Asthma Education and Prevention Program (NAEPP), National Institutes of Health. August 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Full Report 2007. NIH Publication No. 07-4051.
2.Sumino K, Cabana MD. Medication adherence in asthma patients. Curr Opin Pulm Med 2013;19:49–53.
3.Finkelstein JA, Lozano P, Shulruff R, Inui TS, Soumerai SB, Ng M, Weiss KB. Self-reported physician practices for children with asthma: are national guidelines followed? Pediatrics 2000;106: 886–896.
4.Colice GL, Ostrom NK, Geller DE, et al. The CHOICE survey: high rates of persistent and uncontrolled asthma in the United States. Annals of Allergy Asthma & Immunology. 2012 Mar.108(3) 157-U116.
5.Slejko, Julia F., Vahram H. Ghushchyan, Brandon Sucher, Denise R. Globe, Shao-Lee Lin, Gary Globe, and Patrick W. Sullivan. Asthma Control in the United States, 2008-2010: Indicators of Poor Asthma Control. Journal of Allergy and Clinical Immunology 133.6 (2014): 1579-587.
6.National Center for Health Statistics. Summary Health Statistics for US Adults: National Health Interview Survey, 2011. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/nchs/fastats/asthma.htm. Accessed December 8, 2015.
7.National Asthma Control Program. Asthma's Impact on the Nation. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/asthma/impacts_nation/default.htm?s_cid=tw_DrCP274. Accessed December 8, 2015.
8.Diagram 4.1. Relationship of Process Improvements to Outcomes. October 2014. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/asthmaqual/asthmacare/diagram4-1.html
9.Peters SP, Anthonisen N, Castro M, et al. Randomized comparison of strategies for reducing treatment in mild persistent asthma. N Engl J Med. 2007;356(20):2027–2039.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The measure has been in use in the Asthma Specialist Tool to Help Manage Asthma and Improve Quality (ASTHMA IQ), an on-line asthma specialist, pulmonary and primary care provider tool designed to educate providers on the asthma guidelines and help them improve the quality of care provided to asthma patients. ASTHMA IQ is approved for part IV Maintenance of Certification and has been in use since 2008. Data for the Asthma: Pharmacologic Therapy for Persistent Asthma measure in ASTHMA IQ was collected from 21 clinic sites across the country and was manually extracted following chart review.

2008 – 2015 ASTHMA IQ Data:
Average Baseline Performance Score: 84%
Average Performance Score Following Baseline Review and Intervention: 93.6%

ASTHMA IQ. 11 Nov. 2015. Raw data. American Academy of Allergy Asthma and Immunology, Milwaukee.

The following information is from the 2013 Physician Quality Reporting System Experience Report. In 2013, the upper age limit of the measure was 50 years old. In 2014, the upper age limit was changed to 64 years old and in 2015, the upper age limit was removed. PQRS data for 2014 and 2015 is not currently available.
PQRS Measure: #53 Asthma Pharmacologic Therapy for Persistent Asthma
Table A16. Submission Information for Individual PQRS Measures submitted through the Claims Mechanism (2013):
Eligible Professionals: 66,724
Professionals reporting who reported >= 1 valid quality data code: 1777
% of Eligible Professionals who reported >= 1 valid quality data code: 2.6%

Eligible Professionals who satisfactorily reported: 1,519
% of Eligible Professionals who satisfactorily reported: 85.50%
Average Reporting Rate per Eligible Professional: 63.10%
Table A27. Reporting and Performance Information by PQRS Individual Measure (2010 to 2013)
Average Percent of Instances Reported in 2013: 82.0%
Average Performance Rate in 2013: 89.4%
Average Performance Rate in 2012: 92.9%
Average Performance Rate in 2011: 69.1%
Percent of Eligible Professionals Who Had a Performance Rate of At Least 90 Percent: 38.7%

US Dept of Health and Human Services. Centers for Medicare & Medicaid Services. Appendix 2013 Reporting Experience Including Trends (2007-2014) Physician Quality Reporting System and Electronic Prescribing (eRx) Incentive Program. Web. 9 Dec. 2015.

This measure was also used in the CMS Physician Quality Reporting Initiative/System (PQRI/S) in the 2007 through 2011 claims option; 2009 through 2011 registry option; and the 2011 asthma measure group and group practice reporting II options. From 2007-2010, the measure was reported with an upper age limit of 40. In 2011, the upper age limit was changed to 50 years of age.

There is a gap in care as shown by this 2008 data; 46.29% of patients reported on did not meet the measure.(1)

10th percentile: 0.0%
25th percentile: 0.0%
50th percentile: 100.00%
75th percentile: 100.00%
90th percentile: 100.00%

Exception rate: 17.80%

(1)Confidential CMS PQRI 2008 Performance Information by Measure. Jan-Sept TAP file.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
The CHOICE (Comprehensive Survey of Healthcare Professionals and Asthma Patients Offering Insight on Current Treatment Gaps and Emerging Device Options) survey conducted in 2011 found that almost half (490) of the patients participating in the survey were not using controller medications. Of this population, 70% had persistent asthma and 47% had mild or moderate persistent asthma. The survey was conducted over a 12 month period during which time patients not on asthma controllers reported overnight hospitalizations in 7.1% of cases, 12.9% had an ED or outpatient visit and 17.1% had an unscheduled office visit for asthma. Use of an acute care facility was greatest among patients with a higher severity level of asthma and least utilized in patients with well-controlled asthma. The study concluded that many patients with untreated asthma would benefit from long-term controllers and further noted that better management was needed of patients on long-term controller medications.

Colice GL, Ostrom NK, Geller DE, et al. The CHOICE survey: high rates of persistent and uncontrolled asthma in the United States. Annals of Allergy Asthma & Immunology. 2012 Mar.108(3) 157-U116.

A 2013 study of Medical Expenditure Panel Survey data from 2008-2010 examined the national prevalence of self-reported asthma and trends in medication use. Findings included:
•14.6% of respondents indicated use of 3 or more canisters of quick-relief inhalers in the past 3 months
•Of those using short-acting inhalers, 60% used daily long-term control medication but still depended on the use of short-acting inhalers
•28% had never used a long-term control medication and 54% of respondents who had experienced a recent asthma exacerbation indicated they have never used long-term control medication

Slejko, Julia F., Vahram H. Ghushchyan, Brandon Sucher, Denise R. Globe, Shao-Lee Lin, Gary Globe, and Patrick W. Sullivan. "Asthma Control in the United States, 2008-2010: Indicators of Poor Asthma Control." Journal of Allergy and Clinical Immunology 133.6 (2014): 1579-587.

A study published in 2006 found that from 1998-2000:

•43% of patients who reported using a beta2-agonist inhaler more than three times a day on a daily basis were prescribed a longer acting bronchodilator and/or anti-inflammatory agent.
•60% of patients who required chronic treatment with systemic corticosteroids during any 12 month period were prescribed inhaled corticosteroids during that same time period.
Technical Appendix to McGlynn EA, A. S., Adams JL, et al. (2006). "Who is at greatest risk for receiving poor quality health care?" N Engl J Med 354: 1147-1156.

Another study evaluated quality of care for children in the United States. From 1998-2000:
•43% of patients who reported using a beta2-agonist inhaler more than three times a day on a daily basis were prescribed a longer acting bronchodilator and/or anti-inflammatory agent.
•8% of patients requiring chronic treatment with oral corticosteroids had a trial of inhaled corticosteroids first
•80% of patients who required frequent bursts of prednisone who were not already on inhaled corticosteroids or cromolyn were started on them

Mangione-Smith R, D. A., Setodji CM, et al. (2007). "The quality of ambulatory care delivered to children in the United States: Supplementary Appendix." N Engl J Med 357: 1515-1523.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
In 2012, the Centers for Disease Control and Prevention's National Center for Health Statistics (NCHA) released a summary report that provided national estimates for a board range of health measures for the US population. According to the study:
•Adults in poor families had higher percentages of emphysema, asthma, chronic bronchitis, and COPD than adults in families that were not poor.
•Women were more likely to have been told they had asthma, hay fever, sinusitis, or chronic bronchitis than men
•Among adults under age 65, those insured by Medicaid had higher percentages of emphysema, asthma, chronic bronchitis, and COPD than those with private insurance or who were uninsured.
•Among adults aged 65 and over, those insured by Medicare and Medicaid had higher percentages of asthma and chronic bronchitis than those with only Medicare or those with private insurance, and higher percentages of COPD than those with private insurance

Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. National Center for Health Statistics. Vital Health Stat 10(260). 2014.

Similarly, according to the Centers for Disease Control (CDC), disparities persist in asthma prevalence and mortality.
•Gender: Asthma is more likely in women than men and in children, asthma is more likely in boys than girls
•Race and ethnicity: Multi-race and black adults are more likely to have asthma than white adults, black children are 2 times more likely to have asthma than white children, black Americans are 2 to 3 times more likely to die from asthma than any other racial or ethnic group
•Income level: Adults with an annual household income of $75,000 or less are more likely to have asthma than adults with higher incomes

National Asthma Control Program. Asthma's Impact on the Nation. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/asthma/impacts_nation/default.htm?s_cid=tw_DrCP274. Accessed February 3, 2014.

A 2009 retrospective, population-based cohort study of 9,131 adult Medicaid patients with COPD, asthma, or both conditions sought to explain the higher mortality and morbidity seen among African-American patients with these conditions than for their white counterparts. After controlling for age, gender, cohort allocation, and comorbidities, the study found that African-American adults with COPD, asthma, or coexisting COPD and asthma used fewer medical services and accounted for lower medical costs than white adults. The researchers concluded that lower health services utilization and medical costs among African-American patients with COPD and asthma may provide a possible explanation for the racial disparities in outcomes of patients with these conditions. (Shaya, 2009)

Shaya FT, Maneval MS, Gbarayor CM, et. al. Burden of COPD, asthma, and concomitant COPD and asthma among adults: Racial disparities in a Medicaid population. CHEST 2009; 136:405–411.

A 2011 cohort study with 126,019 participants sought to identify ethnic differences for risk of hospitalization for asthma and COPD. Compared with whites, relative risks for asthma among other groups were: blacks, 1.7; Hispanics, 0.9; and Asians, 1.6. Among Asians, increased risk was concentrated in Filipino men and women and South Asian men. (Tran, 2011)

Tran HN, Siu S, Iribarren C, et. al. Ethnicity and risk of hospitalization for asthma and chronic obstructive pulmonary disease. Ann Epidemiol. 2011 Aug;21(8):615-22.

The above studies indicate that a performance gap persists concerning controller agents for patients with persistent asthma. The high quality of evidence, based on numerous RCTs demonstrating the therapeutic utility of controller agents, implies a high degree of confidence is warranted for compliance with this measure being associated with improve patient care outcomes.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, High resource use

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

An estimated 300 million people worldwide suffer from asthma, and it is estimated that by 2025, the prevalence will grow by more than 100 million (WHO, 2007).

In 2012, 29.6 million US adults had been diagnosed with asthma and nearly 19 million adults stated that they currently had asthma. (Blackwell, 2014) In 2011, an estimated 7.1 million US children had asthma (CDC, 2012)

In 2009, asthma resulted in: 479,300 hospitalizations, 1.9 million emergency department visits and 8.9 million doctor visits (CDC, 2011)

Asthma costs the US an estimated $56 billion in medical costs, lost school and work days, and early deaths (CDC, 2011). Asthma is one of the leading causes for missed school in the pediatric population; resulting in missed work for the parents.

Prescription drugs represented the largest single direct medical expenditure related to asthma, costing over $6 billion (WHO, 2007).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. National Center for Health Statistics. Vital Health Stat 10(260). 2014.

National Center for Environmental Health. Asthma in the US; CDC vital signs, 2011. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2011. http://www.cdc.gov/VitalSigns/Asthma/index.html. Accessed February 3, 2014.

National Center for Health Statistics. Summary Health Statistics for US Adults: National Health Interview Survey, 2011. Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2012. http://www.cdc.gov/nchs/fastats/asthma.htm. Accessed January 17, 2014.

World Health Organization. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach, 2007

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
The updated specifications for this measure are submitted within this form. Additional measure information can be found at http://www.aaaai.org/practice-resources/practice-tools/quality-measures.aspx.

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
In 2015, the denominator of this measure was modified to take out the age range maximum, which was previously set at 64 years old. The current age range is 5 years old and older. The upper age limit had undergone several changes. From 2011-2013, the upper age limit was 50 years old. In 2014, it was modified to 64 years old and subsequently removed in 2015. This change was made to include the increasing population of older asthma patients, an important asthma patient population that should qualify for the measure. The 2013 PQRS Experience report showed that providers were reporting this measure on patients outside of the age range. We anticipate that this change will significantly decrease the number of reported instances where the patient did not meet the age range and increase the applicability of this measure to a greater population of patients, notably Medicare eligible patients (age 65 years old and older) who were inadvertently excluded in previous versions of this measure.

Individual PQRS Measures Submitted through the Claims Mechanism Where More than 20% of Reported Instances were for a Patient Who Did Not Meet the Necessary Age Range (2013:
Measure 0047: Asthma: Pharmacologic Therapy for Persistent Asthma
Number of QDCs Reported: 14,936
Number of QDCs without Necessary Age Range Code: 10,136
Percent of QDCs without Necessary Age Range Code: 67.90%

US Dept of Health and Human Services. Centers for Medicare & Medicaid Services. 2013 PQRS and ERx Incentive Program Release Appendix. Web. 9 Dec. 2015.

The denominator instructions were also updated to provide further clarification in accordance with the EPR-3 asthma guidelines. Previously the denominator stated:
Denominator Instructions: Documentation of persistent asthma must be present. One method of identifying persistent asthma is, at

a minimum, daily use of short-acting bronchodilators
Per the EPR-3 guidelines, these instructions are applicable for patients with mild persistent asthma. This was clarified and additional examples were provided in the updated instructions as follows:
Denominator Instructions: Documentation of persistent asthma must be present. One method of identifying persistent asthma is, at a minimum, more than twice a week but not daily use of short-acting bronchodilators for mild-persistent asthma, daily use for moderate persistent asthma; and several times a day for severe persistent asthma.
Evidence for this change can be found in figure 14, EPR-3 Summary Report.
Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. J Allergy Clin Immunol 2007;120:S94–S138.

Additionally, the numerator of the measure has been updated. New tables containing the generic drug names of the applicable medications were developed for each numerator option: inhaled corticosteroids or alternative long-term control medications. The following changes were made to the numerator:
•Deletion of terms anti-asthmatic combinations, mast cell stabilizers, and methylxanthines
•Replacement of the term “antibody inhibitor” with “asthma biologic agents”
•Table of inhaled corticosteroids which includes: beclomethasone, budesonide, ciclesonide, flunisolide, fluticasone, and mometasone
•Table of alternative long-term control medications which includes:
oInhaled steroid combinations: budesonide-formoterol; fluticasone-salmeterol; fluticasone-vilanterol; mometasone-formoterol
oAsthma biologic agents: mepolizumab; omalizumab
oLeukotriene modifiers: montelukast; zafirlukast; zileuton

Specific changes to the measure listed above were made for the following reasons:
•Anti-asthmatic combinations were deleted because they are either covered under inhaled steroid combinations or contain methylxanthines, which is no longer considered adequate controller therapy as a single agent
•Antibody inhibitor is better described as "asthma biologic agents"
•Mast cell stabilizers have limited availability--nedocromil is no longer available, and cromolyn is only available as a nebulizer solution
•According to the EPR-3 asthma guidelines (National Heart, Blood and Lung Institute (NHLBI), National Asthma Education and Prevention Program (NAEPP), National Institutes of Health. August 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Full Report 2007. NIH Publication No. 07-4051), studies comparing ICS to cromolyn or theophylline are limited, but available evidence shows that neither of these long-term control medications appears to be as effective as ICS in improving asthma outcomes (NHLBI/NAEPP, pg. 234). Cromolyn, although having an excellent safety profile, require administration four times per day and, as noted above, is now only available as a nebulizer solution. Theophylline is less desirable because of its safety profile and the need to adjust dose based on diet, drug interactions, and variable metabolism with age (NHLBI/NAEPP, pg. 301). The main use of theophylline is as adjunctive therapy (NHLBI/NAEPP, pg. 234).

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed long-term control medication

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once during the measurement period

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed long-term control medication

Definition:
Long-Term Control Medication Includes: Patients prescribed inhaled corticosteroids (the preferred long-term control medication at any step of asthma pharmacological therapy)

OR
Patients prescribed alternative long-term control medications (inhaled steroid combinations, asthma biologic agents, leukotriene modifiers)

Prescribed: May include prescription given to the patient for inhaled corticosteroid OR an acceptable alternative long-term control medication at one or more visits in the 12-month period OR patient already taking inhaled corticosteroid OR an acceptable alternative long-term control medication as documented in current medication list.

Table 1: Preferred Asthma Control Medication - Inhaled Corticosteroids
beclomethasone
budesonide
ciclesonide
flunisolide
fluticasone
mometasone

Table 2: Alternative Long-term Control Medications
Inhaled steroid combinations: budesonide-formoterol; fluticasone-salmeterol; fluticasone-vilanterol; mometasone-formoterol
Asthma biologic agents: mepolizumab; omalizumab
Leukotriene modifiers: montelukast; zafirlukast; zileuton

For Claims:
Report CPT Category II code:
Performance Met: Inhaled corticosteroids prescribed (4140F)
OR
Performance Met: Alternative long-term control medication prescribed (4144F)
OR
Patient Performance Exclusion: Documentation of patient reason(s) for not prescribing inhaled corticosteroids or alternative long-term control medication (eg, patient declined, other patient reason) (4140F with 2P)
OR
Performance Not Met: Inhaled corticosteroids or alternative long-term control medication not prescribed, reason not otherwise specified (4140F with 8P)

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients aged 5 years and older with a diagnosis of persistent asthma

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Maternal Health, Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
All patients aged 5 years and older with a diagnosis of persistent asthma

Denominator Instructions: Documentation of persistent asthma must be present. One method of identifying persistent asthma is, at a minimum, more than twice a week but not daily use of short-acting bronchodilators for mild-persistent asthma, daily use for moderate persistent asthma; and several times a day for severe persistent asthma.

Denominator Criteria (Eligible Cases):
Patients aged = 5 years on date of encounter
AND
Diagnosis for asthma (ICD-10-CM): J45.30, J45.31, J45.32, J45.40, J45.41, J45.42, J45.50, J45.51, J45.52, J45.901, J45.902, J45.909, J45.990, J45.991, J45.998
AND
Patient encounter during the reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99341,

99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350
AND
Persistent Asthma (mild, moderate or severe): 1038F
**Note: If ICD-10 CM codes J45.30-J45.52 are used to identify the denominator, CPT II code for 1038F is not required; these ICD-10 CM codes capture "persistent asthma".

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator Exceptions:
Documentation of patient reason(s) for not prescribing inhaled corticosteroids or alternative long-term control medication (eg, patient declined, other patient reason)

The AAAAI follows PCPI exception methodology and PCPI distinguishes between measure exceptions and measure exclusions. Exclusions arise when patients who are included in the initial patient or eligible population for a measure do not meet the denominator criteria specific to the intervention required by the numerator. Exclusions are absolute and apply to all patients and therefore are not part of clinical judgment within a measure.

For this measure, exceptions may include patient reason(s) (eg, patient declined). Although this methodology does not require the external reporting of more detailed exception data, the AAAAI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. In further accordance with PCPI exception methodology, the AAAAI advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
For Claims:
Report CPT Category II code with modifier:
4140F-2P: Documentation of patient reason(s) for not prescribing inhaled corticosteroids or alternative long-term control medication (eg, patient declined, other patient reason)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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.)*
To calculate performance rates:
1) Find the patients who meet the initial patient population (ie, the general group of patients that the performance measure is designed to address).
2) From the patients within the initial patient population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial patient population and denominator are identical.
3) From the patients within the denominator, find the patients who qualify for the numerator (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.
4) From the patients who did not meet the numerator criteria, determine if the physician has documented that the patient meets any criteria for denominator exception when exceptions have been specified. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. –Although exception cases are removed from the denominator population for the performance calculation, the number of patients with valid exceptions should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure does not require sampling or a survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Registry, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not Applicable

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)* Ambulatory Care : Clinician Office/Clinic If other:
S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
2a. Reliability – See attached Measure Testing Submission Form **2b. Validity – See attached Measure Testing Submission Form** 0047_MeasureTesting_MSF5.0_Data.doc,0047_Testing_Attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0047
Measure Title: Asthma: Pharmacologic Therapy for Persistent Asthma
Date of Submission: 2/1/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses

whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Initial Testing:

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

Additional Testing:

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*). N/A N/A

1.3. What are the dates of the data used in testing?

Initial Testing: Data collected from patients seen between 01/01/2011-12/31/2011.Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

Additional Testing: 01/01/2014-12/31/2014

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Initial Testing:

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

Additional Testing:

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Initial Testing:
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 86 patient encounters.

Additional Testing:
This project sought to successfully collect de-identified patient-level data from Asthma, Allergy and Immunology practices through the Allergy Partner's central Electronic Medical Record (EMR) system to report accurate denominator and numerator information for the Asthma: Pharmacologic Therapy for Persistent Asthma measure. In order to achieve a reliable sample of patients for this measure, AAAAI and Allergy Partners sought a minimum of 1,000 combined patient records from the across the Asthma, Allergy and Immunology sites that are part of the Allergy Partners network.

The Team worked together to identify sites that were part of the Allergy Partners network that served patients included in the eligible patient population (denominator) for this Measure. The Allergy Partners staff has access to all the data from its 45 practices at 106 different locations which are included in one central EMR system. They sought to identify geographically diverse sites and diverse practice types (i.e. solo vs. group practices) to be included in the random patient sample.

SEA developed a data collection guide and detailed file specifications to educate and assist the Allergy Partners staff in pulling the data from the EMR system. As part of participating in the data collection process, SEA worked with Allergy Partners to go through a denominator certification process to ensure that they were using the appropriate measure parameters and collecting/pulling data from the EMR in a standardized way.

Once the denominator certification process was complete, Allergy Partners randomly chose 44 sites that include patients eligible for the measure (denominator) and that represented geographically diverse areas and with practice type diversity (solo vs. group practice). The de-identified data was sent securely to the full Team for their review. The Statisticians and SEA performed quality checks on the data. Clarifications were made to the age in the denominator (i.e. age at visit ≥5 years old rather than at start of measurement period).

There was not adequate time to do construct validity testing given that only 30 days was allowed to submit the final testing report and given that AAAAI did not have time to request IRB approval to look at non de-identified patient charts to assess construct validity with an authoritative source. In lieu of construct validity, the Team assessed the Measure's Face Validity. Through a systematic and transparent process, by identified experts, the AAAAI sought to explicitly address whether this Measure's performance scores as specified could be used to distinguish good from poor quality. The AAAAI contacted well known experts in Asthma who were not part of the Measure development group for this measure to complete the face validity survey. The results of the survey can be found in Validity section of this report.

The data set included data from 1,863 patients from 44 clinic sites.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Initial Testing:
Measure testing was conducted by AMA-PCPI.

Additional Testing:
The data set included data from 1,863 patients from 44 clinic sites.

Descriptive Measure Statistics

Table 1: Patient Payer Information

	PAYER				
	Commercial	**Medicaid**	**Medicare**	**Self/Uninsured**	**Total**
Clinic A in CA	2	0	0	0	2
Clinic AA in SC	27	0	6	0	33
Clinic B in CA	121	0	14	4	139
Clinic BB in TN	4	0	0	0	4
Clinic C in CA	11	0	2	0	13
Clinic CC in TN	1	0	0	0	1
Clinic D in GA	64	4	10	1	79
Clinic DD in TN	1	0	0	0	1
Clinic E in GA	1	0	0	0	1
Clinic EE in TN	10	0	1	1	12
Clinic F in GA	1	0	0	0	1
Clinic FF in TN	9	0	2	1	12
Clinic G in IN	25	0	6	0	31
Clinic GG in TN	14	0	3	0	17
Clinic H in IN	13	1	4	0	18
Clinic HH in TX	27	0	2	2	31
Clinic II in TX	52	1	18	1	72
Clinic J in IN	7	0	2	0	9
Clinic JJ in TX	54	0	10	3	67
Clinic K in IN	8	0	1	0	9
Clinic KK in TX	27	0	2	0	29
Clinic L in KY	4	0	0	0	4
Clinic LL in VA	66	0	3	0	69
Clinic M in MO	229	10	49	5	293
Clinic MM in VA	2	0	0	0	2
Clinic N in NC	3	0	1	0	4
Clinic NN in VA	1	0	0	0	1
Clinic O in NC	35	22	3	0	60
Clinic OO in VA	1	0	0	0	1
Clinic P in NC	3	0	0	0	3

Clinic PP in VA	151	3	30	2	186
Clinic Q in NC	76	51	23	1	151
Clinic QQ in VA	148	1	43	1	193
Clinic R in NC	8	0	0	0	8
Clinic RR in VA	145	0	26	2	173
Clinic S in NC	1	1	0	0	2
Clinic SS in VA	4	0	1	0	5
Clinic T in NC	2	0	0	0	2
Clinic U in NC	16	11	5	0	32
Clinic V in NM	1	0	0	0	1
Clinic W in NM	21	0	0	0	21
Clinic X in NM	1	0	0	0	1
Clinic Y in NM	52	0	14	3	69
Clinic Z in NV	1	0	0	0	1
Total	1450	105	281	27	1863

Table 2: Patient Place of Residence (based on patient zip code)

	STATE														
	CA	**GA**	**IL**	**IN**	**KY**	**MO**	**NC**	**NM**	**NV**	**NY**	**SC**	**TN**	**TX**	**VA**	**Tot**
Clinic A in CA	2	0	0	0	0	0	0	0	0	0	0	0	0	0	
Clinic AA in SC	0	1	0	0	0	0	0	0	0	0	32	0	0	0	3
Clinic B in CA	138	0	0	0	0	0	0	0	0	0	0	0	1	0	13
Clinic BB in TN	0	0	0	0	0	0	0	0	0	0	0	4	0	0	
Clinic C in CA	13	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Clinic CC in TN	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
Clinic D in GA	0	79	0	0	0	0	0	0	0	0	0	0	0	0	7
Clinic DD in TN	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
Clinic E in GA	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
Clinic EE in TN	0	0	0	0	0	0	0	0	0	0	0	12	0	0	1
Clinic F in GA	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
Clinic FF in TN	0	0	0	0	1	0	0	0	0	0	0	11	0	0	1
Clinic G in IN	0	0	0	31	0	0	0	0	0	0	0	0	0	0	3
Clinic GG in TN	0	0	0	0	0	0	0	0	0	0	0	17	0	0	1
Clinic H in IN	0	0	0	18	0	0	0	0	0	0	0	0	0	0	1
Clinic HH in TX	0	0	0	0	0	0	0	0	0	0	0	0	31	0	3
Clinic II in TX	0	0	0	0	0	0	0	7	0	0	0	0	65	0	7
Clinic J in IN	0	0	0	9	0	0	0	0	0	0	0	0	0	0	
Clinic JJ in TX	0	0	0	0	0	0	0	0	0	0	0	0	67	0	6
Clinic K in IN	0	0	0	9	0	0	0	0	0	0	0	0	0	0	
Clinic KK in TX	0	0	0	0	0	0	0	0	0	0	0	0	29	0	2
Clinic L in KY	0	0	0	0	3	0	0	0	0	1	0	0	0	0	
Clinic LL in VA	1	0	0	0	0	0	1	0	0	0	1	0	0	66	6
Clinic M in MO	0	0	13	0	1	279	0	0	0	0	0	0	0	0	29
Clinic MM in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	2	

Clinic N in NC	0	0	0	0	0	0	4	0	0	0	0	0	0	0	
Clinic NN in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Clinic O in NC	0	0	0	0	0	0	60	0	0	0	0	0	0	0	6
Clinic OO in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Clinic P in NC	0	0	0	0	0	0	3	0	0	0	0	0	0	0	
Clinic PP in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	186	18
Clinic Q in NC	0	0	0	0	0	0	151	0	0	0	0	0	0	0	15
Clinic QQ in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	193	19
Clinic R in NC	0	0	0	0	0	0	8	0	0	0	0	0	0	0	
Clinic RR in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	173	17
Clinic S in NC	0	0	0	0	0	0	2	0	0	0	0	0	0	0	
Clinic SS in VA	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
Clinic T in NC	0	0	0	0	0	0	2	0	0	0	0	0	0	0	
Clinic U in NC	0	0	0	0	0	0	32	0	0	0	0	0	0	0	3
Clinic V in NM	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
Clinic W in NM	0	0	0	0	0	0	0	21	0	0	0	0	0	0	2
Clinic X in NM	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
Clinic Y in NM	0	0	0	0	0	0	0	69	0	0	0	0	0	0	6
Clinic Z in NV	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
Total	154	82	13	67	5	279	263	99	1	1	33	46	193	627	180

Table 3: Patient Race and Ethnicity Information

	ETHNICITY									
	American Indian/Alaskan Native	**Asian**	**Black /African American**	**Hispanic / Latino**	**Multiracial**	**Other**	**Patient Declined**	**Unspecified**	**White**	**T**
Clinic A in CA	0	0	0	1	0	0	0	0	1	
Clinic AA in SC	0	0	14	0	0	0	0	5	14	
Clinic B in CA	0	11	5	37	1	3	0	35	47	
Clinic BB in TN	0	0	1	0	0	0	0	0	3	
Clinic C in CA	0	0	0	0	0	0	2	11	0	
Clinic CC in TN	0	0	0	0	0	0	0	0	1	
Clinic D in GA	0	0	28	7	2	0	0	3	39	
Clinic DD in TN	0	0	0	0	0	0	0	0	1	
Clinic E in GA	0	0	0	0	0	0	0	0	1	
Clinic EE in TN	0	1	4	0	0	0	0	0	7	
Clinic F in GA	0	0	0	0	0	0	0	0	1	
Clinic FF in TN	1	0	5	0	0	0	0	4	2	
Clinic G in IN	0	0	0	0	0	0	0	9	22	
Clinic GG in TN	0	0	0	1	0	0	0	0	16	
Clinic H in IN	0	0	5	2	0	0	0	2	9	
Clinic HH in TX	0	0	2	16	0	0	0	3	10	
Clinic II in TX	0	1	5	45	0	0	0	3	18	

Clinic J in IN	0	0	1	0	0	0	0	2	6	
Clinic JJ in TX	0	5	10	9	0	0	0	2	41	
Clinic K in IN	0	0	0	0	0	0	0	0	9	
Clinic KK in TX	1	2	2	3	0	0	0	0	21	
Clinic L in KY	0	0	0	0	0	0	0	0	4	
Clinic LL in VA	0	6	5	8	0	0	0	2	48	
Clinic M in MO	0	2	10	2	0	0	0	1	278	
Clinic MM in VA	0	0	0	0	0	0	0	0	2	
Clinic N in NC	0	0	0	0	0	0	0	1	3	
Clinic NN in VA	0	0	0	0	0	0	0	0	1	
Clinic O in NC	0	0	7	2	2	0	0	2	47	
Clinic OO in VA	0	0	0	0	0	0	0	0	1	
Clinic P in NC	0	0	0	0	0	0	0	0	3	
Clinic PP in VA	1	0	49	2	0	0	0	9	125	
Clinic Q in NC	0	0	23	3	3	0	0	7	115	
Clinic QQ in VA	0	2	30	5	1	1	0	11	143	
Clinic R in NC	0	0	5	0	0	0	0	0	3	
Clinic RR in VA	0	3	39	1	1	1	0	12	116	
Clinic S in NC	0	0	2	0	0	0	0	0	0	
Clinic SS in VA	0	0	1	0	0	0	0	0	4	
Clinic T in NC	0	0	1	0	0	0	0	0	1	
Clinic U in NC	0	0	9	3	1	0	0	1	18	
Clinic V in NM	0	0	0	0	0	0	0	0	1	
Clinic W in NM	0	0	0	6	0	0	0	0	15	
Clinic X in NM	0	0	0	0	0	0	0	0	1	
Clinic Y in NM	0	0	0	31	0	0	0	0	38	
Clinic Z in NV	0	0	0	0	0	0	0	0	1	
Total	3	33	263	184	11	5	2	125	1237	

Table 4: Patient Age Breakdown

		AGE							
	5 – 11	**12 - 17**	**18-25**	**26-34**	**35-44**	**45-54**	**55-64**	**65+**	**Total**
Clinic A in CA	0	0	0	0	1	0	1	0	2
Clinic AA in SC	6	1	1	4	2	11	3	5	33
Clinic B in CA	33	12	5	9	18	31	14	17	139
Clinic BB in TN	0	0	0	0	1	2	1	0	4
Clinic C in CA	2	0	1	1	1	2	4	2	13
Clinic CC in TN	0	0	0	1	0	0	0	0	1
Clinic D in GA	23	5	7	1	6	28	3	6	79
Clinic DD in TN	1	0	0	0	0	0	0	0	1
Clinic E in GA	1	0	0	0	0	0	0	0	1
Clinic EE in TN	2	2	1	1	1	4	0	1	12
Clinic F in GA	0	0	0	0	0	1	0	0	1
Clinic FF in TN	0	3	2	3	0	2	1	1	12

Clinic G in IN	5	0	2	3	8	3	4	6	31
Clinic GG in TN	3	0	0	1	0	7	2	4	17
Clinic H in IN	1	0	3	2	2	5	3	2	18
Clinic HH in TX	4	0	2	3	5	13	3	1	31
Clinic II in TX	8	1	5	6	5	17	11	19	72
Clinic J in IN	1	0	2	2	2	1	1	0	9
Clinic JJ in TX	7	3	2	6	16	12	13	8	67
Clinic K in IN	2	1	0	1	0	3	1	1	9
Clinic KK in TX	6	3	3	7	6	1	2	1	29
Clinic L in KY	0	0	0	0	0	3	1	0	4
Clinic LL in VA	12	7	7	8	10	19	3	3	69
Clinic M in MO	54	16	15	30	37	51	45	45	293
Clinic MM in VA	1	1	0	0	0	0	0	0	2
Clinic N in NC	0	1	2	0	0	0	0	1	4
Clinic NN in VA	0	0	1	0	0	0	0	0	1
Clinic O in NC	16	2	3	6	3	24	4	2	60
Clinic OO in VA	0	0	0	1	0	0	0	0	1
Clinic P in NC	2	0	1	0	0	0	0	0	3
Clinic PP in VA	30	8	9	21	30	38	24	26	186
Clinic Q in NC	30	12	5	14	9	47	18	16	151
Clinic QQ in VA	17	14	11	16	27	44	26	38	193
Clinic R in NC	1	0	2	1	2	1	1	0	8
Clinic RR in VA	24	5	14	17	22	38	26	27	173
Clinic S in NC	0	0	0	0	0	2	0	0	2
Clinic SS in VA	1	1	0	0	2	0	0	1	5
Clinic T in NC	0	0	1	0	0	1	0	0	2
Clinic U in NC	8	4	1	1	1	12	1	4	32
Clinic V in NM	0	1	0	0	0	0	0	0	1
Clinic W in NM	0	0	0	1	7	4	6	3	21
Clinic X in NM	0	0	0	0	1	0	0	0	1
Clinic Y in NM	6	6	1	3	13	8	22	10	69
Clinic Z in NV	0	0	1	0	0	0	0	0	1
Total	307	109	110	170	238	435	244	250	1863

Measure Results

	Number of Providers	Number of Patients Submitted	Number of Patients with valid exceptions	Denominator	Number of Patients Prescribed ICS as a long-term(asthma) control med	Percentage (%) of Patients Prescribed ICS as a long-term(asthma) control med	Number of Patients Prescribed Non-ICS as a long-term(asthma) control med	Percentage (%) of Patients Prescribed Non-ICS as a long-term(asthma) control med	Number of Patients Prescribed a long-term(asthma) control med	Percentage of Patien Prescribed long-term(asthm control m
Clinic A in CA	1	2	0	2	2	100	2	100	2	100
Clinic AA in SC	1	33	0	33	32	96.97	32	96.97	33	100
Clinic B in CA	2	139	0	139	111	79.86	91	65.47	138	99.28
Clinic BB in TN	1	4	0	4	4	100	4	100	4	100
Clinic C in CA	2	13	0	13	10	76.92	8	61.54	12	92.31
Clinic CC in TN	1	1	0	1	0	0	1	100	1	100
Clinic D in GA	2	79	0	79	74	93.67	65	82.28	79	100
Clinic DD in TN	1	1	0	1	1	100	1	100	1	100
Clinic E in GA	1	1	0	1	1	100	0	0	1	100
Clinic EE in TN	1	12	0	12	11	91.67	8	66.67	12	100
Clinic F in GA	1	1	0	1	1	100	1	100	1	100
Clinic FF in TN	3	12	0	12	9	75	7	58.33	11	91.67
Clinic G in IN	1	31	0	31	28	90.32	27	87.1	31	100
Clinic GG in TN	2	17	0	17	16	94.12	14	82.35	17	100
Clinic H in IN	1	18	0	18	17	94.44	16	88.89	18	100
Clinic HH in TX	1	31	0	31	31	100	22	70.97	31	100
Clinic II in TX	1	72	0	72	69	95.83	44	61.11	72	100
Clinic J in IN	2	9	0	9	9	100	7	77.78	9	100
Clinic JJ in TX	2	67	0	67	66	98.51	44	65.67	67	100
Clinic K in IN	2	9	0	9	9	100	7	77.78	9	100
Clinic KK in TX	1	29	0	29	25	86.21	23	79.31	29	100
Clinic L in KY	1	4	0	4	3	75	1	25	3	75
Clinic LL in VA	3	69	0	69	62	89.86	47	68.12	68	98.55
Clinic M	1	293	0	293	229	78.16	235	80.2	293	100

in MO										
Clinic MM in VA	2	2	0	2	2	100	0	0	2	100
Clinic N in NC	3	4	0	4	3	75	4	100	4	100
Clinic NN in VA	1	1	0	1	1	100	1	100	1	100
Clinic O in NC	1	60	0	60	58	96.67	47	78.33	60	100
Clinic OO in VA	1	1	0	1	1	100	1	100	1	100
Clinic P in NC	2	3	0	3	3	100	3	100	3	100
Clinic PP in VA	6	186	0	186	168	90.32	114	61.29	185	99.46
Clinic Q in NC	1	151	0	151	145	96.03	134	88.74	151	100
Clinic QQ in VA	7	193	0	193	164	84.97	129	66.84	190	98.45
Clinic R in NC	3	8	0	8	8	100	7	87.5	8	100
Clinic RR in VA	7	173	0	173	148	85.55	96	55.49	172	99.42
Clinic S in NC	1	2	0	2	2	100	2	100	2	100
Clinic SS in VA	3	5	0	5	4	80	3	60	5	100
Clinic T in NC	2	2	0	2	2	100	0	0	2	100
Clinic U in NC	2	32	0	32	27	84.38	16	50	29	90.63
Clinic V in NM	1	1	0	1	1	100	1	100	1	100
Clinic W in NM	3	21	0	21	20	95.24	12	57.14	21	100
Clinic X in NM	1	1	0	1	1	100	1	100	1	100
Clinic Y in NM	3	69	0	69	65	94.2	58	84.06	69	100
Clinic Z in NV	1	1	0	1	1	100	1	100	1	100
Total	86	1863	0	1863	1644	88.24	1337	71.77	1850	99.3

Overall, the Inhaled corticoid steroid rate prescribed for long term control of persistent asthma for this Measure 88.24 % for all clinics. The Non-inhaled corticosteroid rate long term control medication for persistent asthma rate was 71.77% for all sites. The total percentage of patients prescribed long-term control medications for persistent asthma was 99.3%, with some overlap of patients being prescribed BOTH inhaled corticosteroids AND non-inhaled corticosteroids.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below. N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Not available.

2a2. RELIABILITY TESTING
***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Initial Testing:
Refer to the validity section for a description of the analytic methods for our EHR testing project

Additional Testing:
Reliability was calculated according to the methods outlined in a technical report prepared by J.L. Adams titled "The Reliability of Provider Profiling: A Tutorial" (RAND Corporation, TR-653-NCQA, 2009 Available at http://www.rand.org/pubs/technical_reports/TR653.html. In this context, reliability represents the ability of a measure to confidently distinguish the performance of one physician from another. There are 3 main drivers of reliability; sample size, differences between physicians, and measurement error." According to this approach, reliability is estimated with a beta-binomial model. The beta-binomial model is appropriate for measuring the reliability of pass/fail measures such as the Measure discussed in this document. There are also computational advantages to using the beta-binomial model, which is based on the beta distribution for the "true" physician scores. The beta distribution is a very flexible distribution on the interval from 0 to 1. The beta-binomial model assumes the physician's score is a binomial random variable conditional on the physician's true value that comes from the beta distribution.

From the RAND 2009 document, Binomial distribution properties will be used for the pass/fail measures in this report. Conceptually, the HLM subtracts known measurement error variances from the overall observed variance of the provider scores to estimate the provider-to-provider variance. To understand some of the special issues in applying reliability to physician profiling, it is useful to add another level of detail to this reliability formula. In particular, a closer look at what is in the error variance can provide insight:

The binomial error is:

$$\sigma^2_{binomial} = \frac{p(1-p)}{n}$$

where p is the passing rate for a physician.

And the reliability is

$$reliability = \frac{\sigma^2_{provider-to-provider}}{\sigma^2_{provider-to-provider} + \sigma^2_{binomial}} = \frac{\sigma^2_{provider-to-provider}}{\sigma^2_{provider-to-provider} + \frac{p(1-p)}{n}}.$$

In this equation, the provider-specific error variance has been rewritten as the average error variance for a single item (average item error) in a physician's score where *n* is the number of items. Here an item would be a single pass/fail event (0/1). This form of the equation makes it easier to see the typical form of the variance of a mean depending on the population variance and the sample size. This formulation also makes it easier to separate the effects of measurement error in the items from the number of items. In many measurement examples other than physician profiling (e.g., functional status questionnaires), this additional detail is less important since all of the respondents have the same *n*. In some important physician profiling problems, the number of items (or patient outcomes) in the physicians' scores can vary widely from physician to physician.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Initial Testing:
Refer to the validity section for the testing results for our EHR testing project

Additional Testing:
Beta Binomial Model

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.e., *what do the results mean and what are the norms for the test conducted?*)

Initial Testing:
Refer to the validity section for the testing results for our EHR testing project

Additional Testing:
Reliability Results

	Number of Providers	Number of Patients Submitted	Denominator	Reliability of ICS	Reliability of Non-ICS	Reliability of All LTC Meds

Clinic A in CA	1	2	2	1.00	1.00	1.00
Clinic AA in SC	1	33	33	0.77	0.93	1.00
Clinic B in CA	2	139	139	0.72	0.88	0.69
Clinic BB in TN	1	4	4	1.00	1.00	1.00
Clinic C in CA	2	13	13	0.18	0.40	0.02
Clinic CC in TN	1	1	1	1.00	1.00	1.00
Clinic D in GA	2	79	79	0.80	0.87	1.00
Clinic DD in TN	1	1	1	1.00	1.00	1.00
Clinic E in GA	1	1	1	1.00	1.00	1.00
Clinic EE in TN	1	12	12	0.32	0.39	1.00
Clinic F in GA	1	1	1	1.00	1.00	1.00
Clinic FF in TN	3	12	12	0.16	0.37	0.02
Clinic G in IN	1	31	31	0.51	0.77	1.00
Clinic GG in TN	2	17	17	0.48	0.58	1.00
Clinic H in IN	1	18	18	0.51	0.68	1.00
Clinic HH in TX	1	31	31	1.00	0.64	1.00
Clinic II in TX	1	72	72	0.84	0.78	1.00
Clinic J in IN	2	9	9	1.00	0.38	1.00
Clinic JJ in TX	2	67	67	0.93	0.78	1.00
Clinic K in IN	2	9	9	1.00	0.38	1.00
Clinic KK in TX	1	29	29	0.42	0.68	1.00
Clinic L in KY	1	4	4	0.06	0.20	0.00
Clinic LL in VA	3	69	69	0.69	0.79	0.36
Clinic M in MO	1	293	293	0.84	0.96	1.00
Clinic MM in VA	2	2	2	1.00	1.00	1.00
Clinic N in NC	3	4	4	0.06	1.00	1.00
Clinic NN in VA	1	1	1	1.00	1.00	1.00
Clinic O in NC	1	60	60	0.85	0.81	1.00
Clinic OO in VA	1	1	1	1.00	1.00	1.00
Clinic P in NC	2	3	3	1.00	1.00	1.00
Clinic PP in VA	6	186	186	0.86	0.90	0.80
Clinic Q in NC	1	151	151	0.92	0.95	1.00
Clinic QQ in VA	7	193	193	0.82	0.91	0.59
Clinic R in NC	3	8	8	1.00	0.47	1.00
Clinic RR in VA	7	173	173	0.81	0.89	0.78

Clinic S in NC	1	2	2	1.00	1.00	1.00
Clinic SS in VA	3	5	5	0.09	0.20	1.00
Clinic T in NC	2	2	2	1.00	1.00	1.00
Clinic U in NC	2	32	32	0.42	0.60	0.04
Clinic V in NM	1	1	1	1.00	1.00	1.00
Clinic W in NM	3	21	21	0.58	0.50	1.00
Clinic X in NM	1	1	1	1.00	1.00	1.00
Clinic Y in NM	3	69	69	0.79	0.86	1.00
Clinic Z in NV	1	1	1	1.00	1.00	1.00
Total	86	1863	1863	0.98	0.99	0.97

SUMMARY OF RESULTS

1. Inhaled Corticosteroid Long-Term control medication Reliability per Beta Binomial Model **0.98**
2. Non-Inhaled Corticosteroid Long-term Control Medication Reliability per Beta Binomial Model **0.99**
3. Combined Long-Term Control Medication Reliability per Beta Binomial Model **0.97**

There was very small provider to provider or in this Measure clinic to clinic variability. Clinic specific reliability is consistently greater than 0.90, and thus can be considered to be very good. Reliability scores vary from 0.0 to 1.0, with a score of 0 indicating that all variation is attributable to measurement error (noise or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across accountable entities. **Values above 0.7 are considered to be sufficient to see differences between some physicians or clinicians and the mean and values above 0.9 are considered sufficient to see differences between pairs of clinics (or physicians) (see the RAND tutorial 2009).**

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

<u>Initial Testing:</u>

EHR Measure Validity

Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:

• Percent agreement at the denominator and numerator (exception - for those measures with exception)

• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

Face Validity

Face validity of the measure score as an indicator of quality was systematically assessed as follows.

After the measure was fully specified, the expert panel (workgroup membership) was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Scale 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5= Strongly Agree

The expert panel consists of 14 members of a Joint Task Force on Quality and Performance Measures convened by the American Academy of Allergy, Asthma & Immunology (AAAAI) and the American College of Allergy, Asthma & Immunology (ACAAI), and facilitated by staff of the AAAAI. Members of this Joint Task Force are all certified by the American Board of Allergy and Immunology. Practice settings of panel members vary significantly and include very small practices, independent large group practices, a large group of affiliated specialty practices, and large academic medical centers.

Michael Schatz, MD MS, FAAAAI (AAAAI Co-Chair), San Diego, CA

Michael Blaiss, MD FAAAAI (ACAAI Co-Chair), Memphis, TN
David Brown, MD, Skyland, NC
Mark Corbett, MD FAAAAI, Louisville, KY
George Green, MD FAAAAI, Abington, PA
David Lang, MD FAAAAI, Cleveland, OH
Eli Meltzer, MD FAAAAI, San Diego, CA
Robert Nathan, MD FAAAAI, Colorado Springs, CO
John Oppenheimer, MD FAAAAI, Cedar Knolls, NJ
Gary Rachelefsky, MD FAAAAI, Los Angeles, CA
Raymond Slavin, MD FAAAAI, St. Louis, MO
Stephen Tilles, MD FAAAAI, Seattle, WA
Dana Wallace, MD FAAAAI, Fort Lauderdale, FL
Robert Wood, MD FAAAAI, Baltimore, MD

Additional Testing:

Validation Results

SEA completed validation of the data in three step process: 1) denominator certification, 2) data file quality checks and 3) Face Validity Survey. Details of this validation are described in this report.

Denominator certification

In order to ascertain that valid and accurate data is being obtained denominator certification is a critical step. The certification required that the data collection body, Allergy Partners, attest that they will submit accurate data and follow the measure specifications precisely as they are written when extracting the data for the Measure from their EMR system from their partner site's data. It was also used to ensure that Allergy Partners was using the appropriate codes and ranges to identify eligible patients for this Measure:

- **Diagnosis Codes for Asthma (ICD-9 as ICD-10 did not go into effect until 2015):** 493.00, 493.01, 493.02, 493.10, 493.11, 493.12, 493.20, 493.21, 493.22, 493.81, 493.82, 493.90, 493.91, 493.92
- **Patient encounter during the reporting period (CPT):** 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350
  AND
- **CPT-II Modifier for Persistent Asthma (mild, moderate or severe):** 1038F
- **Date of birth (≥5 years old at date of the visit; 01/01/1900-01/01/2009)**
- **Date of service range (i.e. 01/01/2014-12/31/2014)**

SEA did not identify any major errors during the review of the denominator information with Allergy Partners and therefore Allergy Partners passed the denominator certification. There were a few corrections and clarifications that required communication between SEA and Allergy Partners; however, each issue was resolved in a timely manner. These issues included:

- Clarifying the minimum age required at time of visit
- Clarifying the diagnosis codes to be included in the data query

Data File Quality Checks

After the data file was submitted by Allergy Partners to SEA and the Statisticians on the Team, quality checks of the files were completed. Each column in the data file represented a field of data for each patient row; the following checks were completed:

- Number of patients/rows were reasonable/expected
- Necessary data fields (columns) were included and completed appropriately
- Patient age met the expected age range
- Patient states were primarily within the state of the clinic where they were seen or within the bordering states as expected (missing data requested and provided)
- Race field(s) were included and populated appropriately
- Provided NPI field was included and number of providers was expected (missing data requested and provided)
- Insurance information was included and was reasonable (missing data requested and provided)
- Encounter visit date during the measurement period
- Diagnoses were included and spanned the list of expected diagnosis codes
- Patient reasons for not prescribing a long term control medication for persistent asthma were applied correctly-not able to identify if a patient refused a LTC for asthma.

Missing data noticed in the initial data files (NPI number, payer, patient location) were completed and resent to the Team for review. Other mentionable items include:

- Given the short amount of time and limitations from using a central EMR, the Team did not have time to manually look through each patient file to look for any patient exceptions for this measure. Allergy Partners ran several queries looking for patient exceptions in the free text in the physician notes, the impressions and other parts of the EMR system for the patients included in the data set. They were unable to locate any evidence of the use of a patient exception for this Measure.

FACE VALIDITY SURVEY

An independent group of 29 asthma, allergy and immunology experts (not involved with the Measure development) were provided with background information on the Measure and were surveyed to assess their agreement with the following statement: "The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality."

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Initial Testing:
EHR Measure Validity

Reliability: N, % Agreement, Kappa
Numerator: 86, 90.1%, 0.00* (-0.6579-0.6579 CI)
Denominator: 86, 94.2%, 0.00* (-0.8507-0.8507 CI)

*This is an example of the limitation of the Kappa statistic. While the agreement can be 90% or greater, if one classification category dominates, kappa can be significantly reduced. (http://www.ajronline.org/cgi/content/full/184/5/1391)

Face Validity
The results of the expert panel rating of the validity statement were as follows: N = 8; Mean rating = 4.875 and 100% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings
1 - 0 (Strongly Disagree)
2 - 0
3 - 0 (Neither Agree nor Disagree)
4 - 1
5 - 7 (Strongly Agree)

Additional Testing:
Face Validity Survey Results
Mean rating = 4.79 (out of 5)

Rating Scale	Number who selected the rating
1 - Disagree	0
2	0
3 – Moderate agreement 3	
4 -	0
5 - Agree 26	
Total	29

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Initial Testing:
Mean rating = 4.875 (out of 5).

Additional Testing:
The mean rating was 4.79 (out of 5). This measure was examined by a group of physician asthma, allergy and immunology experts. Out of the 29 participants, 26 (90%) agreed at the highest level that the scores from the measure as specified would provide an accurate reflection of quality and none disagreed.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Initial Testing:
EHR Measure Validity
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 86 patient encounters.
Data collected from patients seen between 01/01/2011-12/31/2011.
Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

Additional Testing:
The data abstractors searched for the use of exceptions for this measure in the central EMR system. They were unable to locate the use of any exceptions for this measure.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Initial Testing:
Exceptions included a patient reason. Exceptions were analyzed for frequency and variability across providers.

Additional Testing:
0 (zero) exceptions were noted in querying the EMR. Given the limited about of time for testing project there was not enough time to manually pull each file to look at individual patient files to confirm that exceptions had not been used.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Initial Testing:
EHR Measure Validity
Although specifications allowed for documented patient exceptions for the Asthma: Pharmacologic Therapy measure, there were no documented exceptions in this project. All sampled patients were able to be assessed.

Additional Testing:
AAAAI still believes that it is valid that some patients may refuse to be prescribed long term control medication and that physicians participating in a pay for performance program should not penalized for this. Therefore, there is still a need for a patient exception for this measure.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.e., *what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Initial Testing:

CMS Physician Quality Reporting Initiative/System:
The inter-quartile range (IQR) was calculated to determine the variability of performance on the measure

CMS Physician Quality Reporting Initiative/System:
1777 eligible cases were reported on for the 2013 program, the most recent year for which data is available.

The following information is for the 2013 program, the only year for which such data is available.
Clinical Condition and Measure: #53 Pharmacologic Therapy
Eligible Professionals: 66,724
Professionals Reporting: 1777
% Professionals Reporting: 0.91%
# Professionals Reporting >=80% of eligible instances: 329
% Professionals Reporting >=80% of eligible instances: 74.27%

CMS Physician Quality Reporting Initiative/System:
337 cases were reported on for the 2008 program, the most recent year for which data is available.

The following information is for the 2009 program, the only year for which such data is available.
Clinical Condition and Measure: #53 Pharmacologic Therapy
Eligible Professionals: 48,882
Professionals Reporting: 443
% Professionals Reporting: 0.91%
# Professionals Reporting >=80% of eligible instances: 329
% Professionals Reporting >=80% of eligible instances: 74.27%

Additional Testing:
N/A

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Initial Testing:
Scores on this measure: N = 337; Mean = 53.71%
10th percentile: 0.0%
25th percentile: 0.0%
50th percentile: 100.00%
75th percentile: 100.00%
90th percentile: 100.00%

The inter-quartile range (IQR) provides a measure of the dispersion of performance. The IQR is 100.00 and indicates that 50% of physicians have performance on this measure ranging from 0.0% and 100.00% and 10% of physicians have performance rates less than or equal to 0.0%.(1)

(1)Confidential CMS PQRI 2008 Performance Information by Measure. Jan-Sept TAP file.

Additional Testing:
N/A

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*) N/A

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of*

specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Initial Testing:
The measure was calculated using data collected using two different methods of collection:
• Automated EHR report
• Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:
• Percent agreement at the denominator and numerator (exception - for those measures with exception)
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

Additional Testing:
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Initial Testing:
EHR Measure Validity

Reliability: N, % Agreement, Kappa
Numerator: 86, 90.1%, 0.00* (-0.6579-0.6579 CI)
Denominator: 86, 94.2%, 0.00* (-0.8507-0.8507 CI)

*This is an example of the limitation of the Kappa statistic. While the agreement can be 90% or greater, if one classification category dominates, kappa can be significantly reduced. (http://www.ajronline.org/cgi/content/full/184/5/1391)

Additional Testing:
N/A

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Initial Testing:
Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:
• Percent agreement at the denominator and numerator (exception - for those measures with exception)
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

Additional Testing:
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*) This measure is collected with a complete sample; there are no missing data on this measure.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*) This measure is collected with a complete sample; there are no missing data on this measure.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*) This measure is collected with a complete sample; there are no missing data on this measure.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
We believe this measure is valid, reliable, and feasible for implementation.

The AAAAI has identified a large multi-clinic practice willing to provide de-identified measures data and a data analysis team to perform additional testing on this measure if testing data can be submitted at a later date.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Physician Quality Reporting System https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/How_To_Get_Started.html Payment Program Physician Quality Reporting System https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/How_To_Get_Started.html The American Academy of Allergy Asthma and Immunology Quality Clinical Data Registry in Collaboration with CECity: https://www.medconcert.com/content/medconcert/AAAAIQIR/ The American Academy of Allergy Asthma and Immunology Quality Clinical Data Registry in Collaboration with CECity Professional Certification or Recognition Program Asthma Specialist Tool to Help Manage Asthma and Improve Quality (ASTHMA IQ) https://www.asthmaiq.org/ The American Board of Internal Medicine (ABIM) Self-Directed Practice Improvement Module (PIM) http://www.abim.org/maintenance-of-certification/earning-points/practice-assessment/default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) The American Academy of Allergy Asthma and Immunology Quality Clinical Data Registry in Collaboration with CECity https://www.medconcert.com/content/medconcert/AAAAIQIR/

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Name of Program: Physician Quality Reporting System
Sponsor: Centers for Medicare and Medicaid Services
Purpose: According to CMS, PQRS is a quality reporting program that encourages individual eligible professionals (Eps) and group practices to report information on the quality of care to Medicare. The Asthma: Pharmacologic Therapy for Persistent Asthma measure is one of two asthma measures in the program and the only asthma measure in the asthma measures reporting group.
Geographic area: National
Number and Percentage of accountable entities and patients included: According to the 2013 PQRS experience report, in 2013, 641,654 (51 percent) eligible professionals (including those who belonged to group practices that reported under the GPRO, eligible professionals within a Medicare ACO participating under the SSP or Pioneer ACO Model, and eligible professionals participating

through the CPC initiative) participated in PQRS

Name of Program: ASTHMA IQ
Sponsor: The American Academy of Allergy Asthma and Immunology
Purpose: ASTHMA IQ is a web-based tool to help track and manage patients with asthma for the purposes of quality improvement. MOC part IV credit is available upon completion.
Geographic area: National
Number and Percentage of accountable entities and patients included:
Data as of 11/13/2012
Users (Specialist): 935
Users (Primary care): 198
Patients (Specialist): 8342
Patients (Primary care): 2045
Clinics (Specialist): 882
Clinics (Primary care): 170

Name of Program: American Board of Internal Medicine's Self Directed Practice Improvement Module
Sponsor: American Board of Internal Medicine
Purpose: To implement a quality-improvement (QI) plan for physician practices. Available for MOC and CME.
Geographic area: National
Number and Percentage of accountable entities and patients included:

Name of Program: The American Academy of Allergy Asthma and Immunology Quality Clinical Data Registry in Collaboration with CECity (AAAAI QCDR):
Sponsor: The American Academy of Allergy Asthma and Immunology
Purpose: A quality reporting tool for PQRS reporting and to foster performance improvement and improve outcomes in the care of patients with allergies and asthma.
Geographic area: National
Number and Percentage of accountable entities and patients included: The AAAAI launched the AAAAI QCDR in 2014, including the Asthma: Pharmacologic Therapy for Persistent Asthma measure, without an upper age limit of 64 years old. The measure is included in the registry for 2015 year and will also be used in 2016. QCDR data on the measure is not yet available from 2014.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of**

high-quality, efficient healthcare for individuals or populations.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

We are not aware of any unintended consequences related to this measurement.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

1799 : Medication Management for People with Asthma

1800 : Asthma Medication Ratio

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications completely harmonized?

No

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

Measures 0047 is similar to NQF measure 1800 (Asthma Medication Ratio) and measure 1799 (Medication Management for People with Asthma) in regards to the denominator population of patients with persistent asthma. However, the denominators differ with respect to the method by which patients with persistent asthma are identified. For measures 1800 and 1799, persistent asthma is defined from administrative data, while for measure 0047, persistent asthma is defined based on clinical information. Additionally, the denominator for measure 0047 been updated to include asthma patients aged 65 and older, an important population that is not reached by measures 1800 and 1799. The numerator for measure 0047 is similar to the numerator in measure 1799, except that inhaled corticosteroids and alternative controllers are reported separately as well as together. The separate reporting rates required by measure 0047 for inhaled corticosteroids and for alternative long-term control medications will be useful for clinicians to assess and manage the use of the preferred vs. alternative long-term control medications for their patients. The numerator of measure 0047 has also been updated to include current and appropriate alternative long-term control medications. While the inhaled corticosteroids in measure 0047 and 1799 are well harmonized, the alternative long-term controllers differ. Measure 1799 includes nedocromil, methylxanthines and cromolyn, all medications that were reviewed by the AAAAI's measure stewardship committee and removed.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** AAFA_letter_of_support_for_NQF_0047..pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The American Academy of Asthma Allergy and Immunology
Co.2 **Point of Contact:** Sheila, Heitzig, sheitzig@aaaai.org, 414-272-6071-
Co.3 **Measure Developer if different from Measure Steward:** The PCPI Foundation
Co.4 **Point of Contact:** Samantha, Tierney, Samantha.Tierney@ama-assn.org, 312-464-5524-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Ann Fuhlbrigge, MD (Co-Chair) (pulmonology)
William E. Golden, MD, MACP (Co-Chair) (internal medicine)
Michael Cabana, MD (pediatrics)
Carlos Camargo, MD, DrPH (emergency medicine)
Tera Crisalida, PA (physician assistant)
Daniel Dressler, MD, MSc (hospital medicine)
Kurt Elward, MD, MPH (internal medicine)
Len Fromer, MD, FAAFP (family medicine)
Gary N. Gross, MD (allergy)
Michael Hagen, MD (family medicine)
Christine Joseph, PhD (epidemiology)
Allan Lieberthal, MD (pediatrics)
Allan Luskin, MD (allergy)
Harold Nelson, MD (internal medicine)
Sai Nimmagadda, MD (pediatric allergy)
Richard D. O'Connor, MD (allergy)
Mimi Saffer (pediatrics)
Michael Schatz, MD (allergy)

PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2003

Ad.3 **Month and Year of most recent revision:** 11, 2015

Ad.4 **What is your frequency for review/update of this measure?** Coding/Specifications updates annually. Review of measures on a three-year cycle, when feasible.

Ad.5 **When is the next scheduled review/update for this measure?** 2016

Ad.6 **Copyright statement:** This Measure is not clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

This Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measure requires a license agreement between the user and the PCPI® Foundation (PCPI®) or American Academy of Allergy, Asthma and Immunology (AAAAI). Neither AAAAI, nor the American Medical Association (AMA), nor the AMA-convened Physician Consortium for Performance Improvement® (AMA-PCPI), now known as the PCPI, nor their members shall be responsible for any use of the Measure.

The AMA's, AMA-PCPI's and National Committee for Quality Assurance's significant past efforts and contributions to the development and updating of the Measures is acknowledged. AAAAI is solely responsible for the review and enhancement ("Maintenance") of the Measure as of November 23, 2015.

AAAAI encourages use of the Measure by other health care professionals, where appropriate.

THIS MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. AAAAI, the AMA, the PCPI and its members and former members of the AMA-PCPI disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0091
De.2. **Measure Title:** COPD: Spirometry Evaluation
Co.1.1. **Measure Steward:** American Thoracic Society
De.3. **Brief Description of Measure:** Percentage of patients aged 18 years and older with a diagnosis of COPD who had spirometry results documented
1b.1. **Developer Rationale:** Despite major efforts to broadly disseminate the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines and use of COPD performance measures across different specialty societies, COPD remains underdiagnosed and misdiagnosed (Collins et al., 2015; Perez et al., 2011). Although spirometry use has increased, it remains underutilized to confirm airflow obstruction and accurately diagnose COPD (CDC, 2012; Nishi et al., 2013). Studies show proper COPD diagnosis with spirometry is done on just over half of patients in the US and Canada (Boulet et al., 2013; Bourbeau et al., 2008; Collins et al., 2015; Nishi et al., 2013; Perez et al., 2011; Yu et al., 2013) and ranges from 10-48% in the Asia-Pacific region, Africa, eastern Europe, and Latin America (Aisanov et al., 2012). A study of physician-diagnosed COPD patients hospitalized for exacerbations found that 22% of patients did not have COPD upon spirometry testing (Prieto Centurion, et al., 2012).

Treatment of COPD without accurate diagnosis and understanding of true etiology of symptoms results in patients not receiving medication that would improve symptoms and quality of life, prevent exacerbations and reduce costly use of emergency and hospital services while other patients may be exposed to adverse effects of unneeded medication and or delays in true diagnosis and management of another condition increasing overall cost of care (Boulet et al., 2013; Bourbeau et al., 2008; CDC, 2012; Collins et al., 2015; Joo et al., 2011). We believe this measure will continue to increase appropriate spirometry use to assist physicians in the accurate diagnosis and treatment of patients with COPD, improving patient management and reducing total costs of COPD.

Citations:

Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.

Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.

Bourbeau J, Sebaldt RJ, Day A, Bouchard J, Kaplan A, Hernandez P, Rouleau M, et al. Practice patterns in the management of chronic obstructive pulmonary disease in primary practice: the CAGE study. Can Respir J. 2008 Jan-Feb:15(1):13-9.

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.

Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.

Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.

Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.

Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.

Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

S.4. **Numerator Statement:** Patients with documented spirometry results in the medical record (FEV1 and FEV1/FVC)
S.7. **Denominator Statement:** All patients aged 18 years and older with a diagnosis of COPD
S.10. **Denominator Exclusions:** Documentation of medical reason(s) for not documenting and reviewing spirometry results
Documentation of patient reason(s) for not documenting and reviewing spirometry results
Documentation of system reason(s) for not documenting and reviewing spirometry results

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Team

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jul 31, 2012

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2013

- The Committee agreed with developer's assessment of evidence.
    - The citations referenced in the updated guidelines that are specific to spirometry use address its application in the detection of COPD. The measure developer cites excerpts from the 2011 update

specific to this use.
    - There are no articles that are specific to the use of spirometry once a patient is diagnosed with COPD (the population for whom the measure is intended).
    - From the 2011 ACP Guideline update: "In our guideline update, there is no new evidence to support the use of routine periodic spirometry after initiation of therapy to monitor disease status or to modify therapy in symptomatic patients. Improvements in clinical symptoms do not necessarily correlate with spirometric responses to therapy or reduction of long-term decline in FEV1. Spirometry is useful to identify symptomatic patients with airflow obstruction who may benefit from pharmacotherapy. Because of the wide intraindividual variation, the spirometric decline of lung function cannot be used to measure individual long-term response to treatment."
- The Committee agreed the guidelines are clear about when spirometry is indicated to confirm the diagnosis of COPD, and that it is not indicated to monitor treatment.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Exception to evidence: NA

Guidance from the Evidence Algorithm: 1→3 →4→5 (eligible for HIGH rating)

Question for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports a performance gap of 45.7% of patients who did not meet this measure in the 2008 PQRS. Based on the measure specifications, the previous Committee thought it was unclear whether this gap focused on the use of spirometry to confirm a COPD diagnosis or on routine spirometry use. The developer did not provide additional information since the last review.

Disparities

- The developer indicates disparities are identified as an issue in the literature, but results for this measure's ability to detect them were not provided.

Questions for the Committee:

- *Does the gap information provided match the measure specifications?*
- *Is 2008 data sufficient to determine that a gap exists now?*
- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare? Based on the literature, is the developer's rationale for not providing disparities information acceptable?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**Process measure of spirometry in persons diagnosed with COPD.
No change in evidence since prior review (2012). Evidence includes: meta-analysis, provision of quality, quantity and consistency of evidence and description of evidence grading.
ANSWER TO QUESTION: As the evidence was high, there is no need to further discuss the evidence.

**Process measure based on high quality evidence.
**I agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence.
**Evidence has not changed and the measure is part of the GOLD multi-college guidelines.

1b. Performance Gap
Comments:
**The developer reports a performance gap of 45.7% of patients who did not meet this measure in the 2008 PQRS, leading to questions re: was this of already diagnosed COPD or screening spirometry? Per the developers, the gap remains limited by available literature.
QUESTIONS:
o Does the gap information provided match the measure specifications? YES--THE LITERATURE IS WHAT IT IS.
o Is 2008 data sufficient to determine that a gap exists now? YES IF NO UPDATED DATA (NONE SUBMITTED)
o Is there a gap in care that warrants a national performance measure? YES--THIS GAP, EVEN IF PARTLY FROM SCREENING, THEN THIS MEASURE IS NEEDED TO MEET THE GOLD STANDARDS
o Are you aware of evidence that disparities exist in this area of healthcare? Based on the literature, is the developer's rationale for not providing disparities information acceptable? YES--See Koskela et al (e pub 2016) COPD Individual FEV1 Trajectories Can Be Identified from a COPD Cohort. This supports differential trajectories in COPD that support the need for ongoing spirometry after diagnosis. An additional article (albeit a review article) supporting this process measure is Cooke et al. (2012) COPD. 2012 Feb;9(1):73-80. Review: clinical inertia in the management of chronic obstructive pulmonary disease.
Subgroups are not identified.
Process measure--disparities in care in that optimal care may not be provided.
**Performance gap exists... more recent data would be preferred but not necessary.
**The gap information provided matches the measure specifications.
A gap (~30%) appears to continue to exist between those with a diagnosis of COPD and those with spirometry confirmed diagnosis of COPD.
**Marked performance gap by EP

1c. High Priority (previously referred to as High Impact)
Comments:
**NA
**NA

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. **Data source(s):** Administrative claims, Electronic Clinical Data : Registry **Specifications:** • The developer attests the specifications were not updated since the last review. • The numerator for this measure is *patients with documented spirometry results in the medical record (FEV1 and FEV1/FVC)*. The denominator is *all patients aged 18 years and older with a diagnosis of COPD*. The previous Committee expressed concern the numerator specifications were unclear and that since the spirometry test should be performed at least once every 12 months, there could be potential for inappropriate/overuse. • The calculation algorithm is stated in S.18 and appears straightforward. ***Questions for the Committee :*** o *Are all the data elements clearly defined? Are all appropriate codes included?* o *Is the logic or calculation algorithm clear?* o *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The time window indicates a 1-year measurement period, but the developer states that a spirometry at any time counts in the numerator. The Committee noted that the specification for "most recent documentation of spirometry" implies that may be several spirometry tests performed and thus confusion.
- Initial reliability testing relied on data-element level validity testing, which per NQF guidance does not require separate reliability testing. Results from this are discussed in the following section.

Describe any updates to testing:

- Performance measure score-level reliability testing conducted by Mathematica Policy Research in 2012
- Testing and analysis included 11,593,241 Medicare beneficiaries identified on claims associated with 2,064 groups of physicians with at least 25 eligible professionals (EPs) (average of 120 EPs per group). Of these, there were 693 groups of physicians with at least 100 EPs (average of 322 EPs).
- This group represents 30% of medical group practices with 25 or more EPs nationwide.
- Groups were included if they reported at least 20 eligible cases for the measure.
- The groups were distributed across all states, the District of Columbia, Guam, and Puerto Rico.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐**No**

Method(s) of reliability testing:

- Reliability was estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error).

Results of reliability testing:

- The developer reports that although there is no universally agreed-upon minimum reliability threshold, reliability scores in the 0.40–0.70 range are considered moderate, and scores greater than 0.70 are considered high, indicating a group's performance rates would be similar if performance were calculated on the basis of a random sample of the practice's beneficiaries.
- The reliability score for this measure was 0.73 among groups with 25 or more EPs and 0.83 among groups with 100 or more EPs.

Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH)

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☒ **No**

Specification not completely consistent with evidence:

- The developer attests the specifications have not been updated. As noted earlier, the previous Committee thought it was unclear whether the reported use and gap focused on the use of spirometry to confirm a COPD diagnosis or on routine spirometry use. The developer did not provide additional information since the last

review.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Empirical testing was conducted at the data element level
  - The results of the data-element testing (1 site, 123 records) conducted by comparing the automated EHR report to visual inspection of the medical record were:
    - Reliability: N, % Agreement, Kappa
    - Numerator: 123, 86.89%, 0.7281 (0.6086-0.8476 CI)
    - Denominator: 123, 100%, kappa non-calculable (non-calculable CI)
- Face validity also was conducted by AMA-PCPI and was assessed by a 12-member expert panel.
- The previous Committee had several concerns regarding the measure's validity:
  - The denominator captures all patients with a diagnosis of COPD – not just newly diagnosed or suspected diagnosis of COPD – making the measure open to misinterpretation. The Committee recommended adding an exclusion for patients previously diagnosed with COPD.
  - The value of stating the lower end of the age range as 18 years was unclear, given the incidence and prevalence of COPD starts climbing after age 40.
  - Validity testing was limited to one academic medical center.
  - The Committee questioned whether the measure is capturing routine use of spirometry after COPD diagnosis, which is not indicated by the evidence or guidelines.

Describe any updates to validity testing:

- The developer, ATS, conducted new face validity testing.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

☐ **Empirical validity testing of the measure score**

Validity testing method:

- The 12-member ATS Clinical Practice Committee was asked to rate its agreement with the following statement: *The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality*.
- The rating scale used was 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

Validity testing results:

- 91.7% of panelists agreed or strongly agreed this measure can accurately distinguish good and poor quality

Questions for the Committee:

- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Do you agree with the previous Committee's concerns regarding the validity of this measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Exceptions include medical reason(s), patient reason(s) or system reason(s) for not documenting spirometry results. Although this methodology does not require the external reporting of more detailed exception data, the

ATS recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness.

- Exclusion analysis was not conducted on this measure, as requested by NQF.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are the exclusions too vague? Are any patients or patient groups inappropriately excluded from the measure?*
- *Because no exclusion analyses were performed and reported by the developer, is there a threat to validity?*
- *Is the Committee aware of whether the exclusions/exceptions are of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Prior-year benchmarks were computed for the claims-based quality indicators, and none of the measures differed significantly at the 5 percent level from the prior year benchmark. A weighted average (based on eligible cases) of performance for groups with 25 or more EPs serves as the benchmark for all groups of this size, whereas a comparable weighted average among groups with at least 100 EPs forms the benchmark for larger groups (100 or more EPs).
    - The percent of groups different than the benchmark ($p<0.05$) for this measure among groups with 25 or more EPs was 45.6%.
    - The percent of groups different than the benchmark ($p<0.05$) for this measure among groups with 100 or more EPs was 47.1%.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Because no analysis of meaningful differences was provided by the developer, is there a threat to validity?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- Missing data analysis was not conducted on this measure

Question for the Committee:

- *Because no analysis of missing data was provided by the developer, is there a threat to validity?*

Guidance from the Validity Algorithm : 1→2→4→5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Question for the Committee:

o Are the specifications consistent with the evidence? UNCERTAIN/LEAN TOWARDS YES

As the specifications have not been updated. Specifications may be limited by the evidence available.

**Whether spirometry is performed in association with NEW diagnosis or progress evaluation may be an issue, as the latter may not be warranted in certain situations for a patient with mild COPD.

**I concur with the previous Committee who thought it was unclear whether the reported use and gap focused on the use of spirometry to confirm a COPD diagnosis or on routine spirometry use.

**Definitions of data acquisition should probably be detailed better

2a2. Reliability Testing

Comments:

**New validity testing was completed by the developer since last review using both measure score and data element testing against a gold standard. Data element was conducted on 123 records at one center with reliability reported as 86.89% on numerator and 100% on denominator. Face validity testing was completed by the 12 member ATS Clinical

Practice Committee with 91.7% agreement that this process measure can accurately distinguish good and poor quality
Questions for the Committee:
- Do you agree that the score from this measure as specified is an indicator of quality? YES
- Do you agree with the previous Committee's concerns regarding the validity of this measure? YES

**In harmonizing this measure with 0102, which I would favor, the age range should be modified to >40
**Until the issue of the specific use of spirometry is included in the algorithm (diagnosis or routine use), the data sought may not be able to be accurately obtained from this measure.
**This can easily be a good indicator for quality, may need better definitions as mentioned above.

2b2. Validity Testing
Comments:
**2b3 Exclusions are Physician reported. Not tested per NQF
Questions for the Committee:
- Are the exclusions consistent with the evidence? YES--standard of care
- Are the exclusions too vague? Are any patients or patient groups inappropriately excluded from the measure? NO--must be based on physician assessment/professional determination
- Because no exclusion analyses were performed and reported by the developer, is there a threat to validity? NO
- Is the Committee aware of whether the exclusions/exceptions are of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? NO

2b4: NO risk adjustment
2b5: Meaningful difference reported:
- The percent of groups different than the benchmark ($p<0.05$) for this measure among groups with 25 or more EPs was 45.6%.
- The percent of groups different than the benchmark ($p<0.05$) for this measure among groups with 100 or more EPs was 47.1%.

Question for the Committee:
- Does this measure identify meaningful differences about quality? YES
- Because no analysis of meaningful differences was provided by the developer, is there a threat to validity? NO

2b6: Not applicable
2b7: No analysis of missing data provided
Question for the Committee:
- Because no analysis of missing data was provided by the developer, is there a threat to validity? UNCERTAIN, BUT DOUBTFUL

Guidance from the Validity Algorithm : 1→2→4→5 (highest eligible rating is MODERATE)
**For this measure, case mix or SDS adjustment not required.
**Exclusions: If the metric being measured is whether spirometry is being used to confirm COPD diagnosis, then an exclusion for follow-up visits that may or may not include spirometry should be developed.
**No threat to validity. Very simple measure. Improved capture with e-reporting
missing data doesn't threat validity

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Reliability testing was completed at both the measure score and data element levels.
Updated testing completed on performance measure score-level by Mathematics Policy Research (2012) using 11M+ Medicare beneficiaries and 2K+ Physician groups with at least 25 eligible professionals/group (avg 120/group). Of these 693 physician groups had at least 100 professionals (avg 322/group)--these represent 30% of medical group practices nationwide. Physician groups were included with minimum 20 eligible cases. Geographic representation was nationwide (all 50 states) as well as DC, Guam and Puerto Rico.
Reliability estimated as ratio of variation between groups and total variation to evaluate variation from measurement error. Developer reports no minimum reliability threshold with reliability score reported as 0.73 in provider groups of 25+ and 0.83 in provider groups of > 100 (both considered in high reliability range).
Questions for the Committee:
- Do the results demonstrate sufficient reliability so that differences in performance can be identified?

Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH)

Questions for the Committee:
o Do the results demonstrate sufficient reliability so that differences in performance can be identified? YES
**Reliability is sufficient
**Sufficient reliability appears to have been demonstrated.
**Differences can easily be identified but on a very general numerator

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**• All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care.
Questions for the Committee:
o Are the required data elements routinely generated and used during care delivery? YES
o Are the required data elements available in electronic form, e.g., EHR or other electronic sources? YES
**Relies on electronic clinical data
**Feasibility seems reasonable.
**The required data elements SHOULD be routinely generated and available within the EHR or paper chart

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure is planned for integration into the CMS Physician Compare Program. Although Physician Compare has been launched, this measure is not yet included as of 12/14/15.
- This measure has been in use for the CMS PQRS program since 2007.

Improvement results: The developer referenced the 2008 PQRS performance data to present improvement results.

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected

findings during implementation.

Potential harms: The developer did not identify any unintended consequences related to this measure.

Feedback: No feedback provided on QPS. Measure reviewed by MAP for Physician Quality Reporting System (PQRS) in 2012, 2013, and 2014. The measure also was reviewed in Physician Compare and Value-Based Payment Modifier Program in 2014. MAP recommended the developer explore creating a composite of all COPD measures and then link that composite with the COPD resource use measure.

***Questions for the Committee**:*

- *The measure is in use, but do are the performance results from 2008 acceptable to indicate usability?*
- *Are you aware of any implementation issues with this measure?*
- *Can you identify any potential harm associated with the measure? Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**This measure is a currently publicly reported and used in accountability programs. This measure is planned for integration into the CMS Physician Compare Program. Although Physician Compare has been launched, this measure is not yet included as of 12/14/15. This measure has been in use for the CMS PQRS program since 2007. Developer states there were no unexpected findings during implementation.

Questions for the Committee:

o The measure is in use, but do are the performance results from 2008 acceptable to indicate usability? IF THEY ARE THE MOST RECENT DATA THEN YES

o Are you aware of any implementation issues with this measure? NO

o Can you identify any potential harm associated with the measure? NO Do the benefits of the measure outweigh any potential unintended consequences? YES

**YES. Benefit outweighs potential unintended consequences.

**Usability and Use appear to be appropriate.

**Implementation burden is minimal, even if better definitions are considered.

The implementation of this measure appears to be devoid of harm and it may improve pt management

Criterion 5: Related and Competing Measures

Related or competing measures

- 0577 : Use of Spirometry Testing in the Assessment and Diagnosis of COPD

Harmonization

- The previous Pulmonary/Critical Care Committee recommended that measures 0091 and 0577 be fully harmonized in order to continue endorsement. At the time, NCQA and PCPI stated the recommendations to address misalignment in the specifications were due to the different data collection and reporting environments, but they would review the differences with their respective measure advisory expert panels to harmonize, if possible.
- ATS, the developer of this measure, did not address harmonization issues.
- NCQA's response to harmonization in its current submission for 0577: These measures have distinct differences in their denominators and numerators. First, our measure is broader in denominator population, being for all patients age 18 years and older with a diagnosis of COPD, while 0577 is for patients age 40 years and older with a new diagnosis of COPD. Our measure is more consistent with COPD guidelines, which do not state an age to start using a spirometry evaluation; rather, spirometry should be used to assess all adults with

COPD, not just adults with a new diagnosis of COPD. Second, our measure´s numerator is more flexible than 0577, allowing a spirometry evaluation anytime during the measurement period, rather than 0577´s requirement that spirometry be performed within 6 months of a new diagnosis of COPD. Our measure numerator is also specific to spirometry results, requiring both the FEV1/FVC values.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM

NOTE: As instructed, responses from the 2012 comprehensive review evidence form submitted by the AMS-PCPI are changed to black font. Items in red font within a response represent minor edits reflecting 2015 updates to GOLD.

NQF #: 0091 **NQF Project:** Pulmonary Project

1. IMPACT, OPPORTUITY, 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)**

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)***:**

The measure focus is the process of providing a spirometry evaluation to all adults with COPD to assist in proper diagnosis and routine treatment of patients with COPD. This process is directly related to reducing COPD exacerbations and inpatient hospitalizations. Proper diagnosis leads to better COPD treatment, which should lead to less comorbid disease, physical dysfunction, and death from COPD.

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

Clinical Practice Guideline

Systematic review of body of evidence (other than within guideline development)

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)***:**

The evidence cited for this measure is directly related to the usefulness of spirometry evaluation in adults with stable COPD. There are no differences from the measure focus and measure target population.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles*)**:** The quantity of studies reviewed in the ACP/ACCP/ATS/ERS guideline was not stated, but the guideline paper references 62 articles. This guideline is based on a targeted literature update from March 2007 to December 2009 to evaluate the evidence and update the 2007 ACP clinical practice guideline on diagnosis and management of stable COPD.

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):* The ACP/ACCP/ATS/ERS guideline recommendation was graded as a strong recommendation, with moderate-quality evidence. A strong recommendation means that benefits clearly outweigh risks and burden, or risks and burden clearly outweigh benefits. Evidence is considered moderate quality when it is obtained from RCTs with important limitations—for example, biased assessment of the treatment effect, large loss to follow-up, lack of blinding, unexplained heterogeneity (even if it is generated from rigorous RCTs), indirect evidence originating from similar (but not identical) populations of interest, and RCTs with a very small number of participants or observed events. In addition, evidence from well-designed controlled trials without randomization, well designed cohort or case–control analytic studies, and multiple time series with or without intervention are in this category. Moderate-quality evidence also means that further research will probably have an important effect on our confidence in the estimate of effect and may change the estimate. (Amir Qaseem, MD, PhD, MHA; Vincenza Snow, MD; Douglas K. Owens, MD, MS; and Paul Shekelle, MD, PhD, for the Clinical Guidelines Committee of the American College of Physicians. The Development of Clinical Practice Guidelines and Guidance Statements of the American College of Physicians: Summary of Methods. Ann Intern Med. 2010;153:194-199.)

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The ACP/ACCP/ATS/ERS guideline recommendation for spirometry is not consistently recommended for all COPD populations. Rather, the guideline explains that targeted use of spirometry for diagnosis of airflow obstruction is beneficial for patients with respiratory symptoms, particularly dyspnea. Existing evidence does not support the use of spirometry to screen for airflow obstruction in individuals without respiratory symptoms, including those with current or past exposure to risk factors for COPD. Evidence is insufficient to support the use of inhaled therapies in asymptomatic individuals who have spirometric evidence of airflow obstruction, regardless of the presence or absence of risk factors for airflow obstruction. There is no difference in the annual rate of FEV1 decline or prevention of symptoms in these individuals with treatment. No evidence from RCTs supports treating asymptomatic individuals, with or without risk factors for airflow obstruction, who do not have spirometric evidence of airflow obstruction. In addition, evidence does not show any independent benefit of obtaining and providing spirometry results on success rates in smoking cessation. No study evaluated the use of periodic spirometry after initiation of therapy to monitor ongoing disease status or modify therapy. (Qaseem et al, 2011)

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

The ACP/ACCP/ATS/ERS guideline panel included representatives from each of the 4 collaborating organizations, and the resulting guideline represents an official and joint clinical practice guideline from those organizations. The guideline panel communicated via conference calls and e-mails. The members reached agreement and resolved any disagreements

through facilitated discussion. The final recommendations were approved by unanimous vote.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The ACP/ACCP/ATS/ERS guideline panel included representatives from each of the 4 collaborating organizations, and the resulting guideline represents an official and joint clinical practice guideline from those organizations. Potential Conflicts of Interest: Any financial and nonfinancial conflicts of interest of the group members were declared, discussed, and resolved. Dr. Wilt: Grant: American College of Physicians; Payment for manuscript preparation: American College of Physicians. Dr. Hanania: Consultancy: GlaxoSmithKline, Boehringer Ingelheim, Novartis, Pfizer, Sunovion, Pearl, Forest; Grants/grants pending (money to institution): GlaxoSmith-Kline, Boehringer Ingelheim, Novartis, Pfizer, Sunovion; Payment for lectures including service on speakers bureaus: GlaxoSmithKline, Astra-Zeneca, Boehringer Ingelheim, Merck. Dr. Criner: Consultancy: Uptake Medical, PortAero, Pulmonx; Grants/grants pending (money to institution): Aeris Therapeutics, Emphysas Medica. Dr. van der Molen: Consultancy: MSD, AstraZeneca, GlaxoSmithKline, Nycomed; Grants/grants pending (money to institution): AstraZeneca, GlaxoSmithKline, Novartis; Payment for lectures including service on speakers bureaus: AstraZeneca, Nycomed, GlaxoSmithKline, MSD. Dr. Marciniuk: Board membership: American College of Chest Physicians, Chest Foundation, Lung Association of Saskatchewan, Canadian COPD Alliance, Canadian Thoracic Society; Consultancy (no payment received): Public Health Agency of Canada, Canadian Agency for Drugs and Technology in Health; Consultancy: Saskatchewan Medical Association; Consultancy (money to institution): AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Saskatchewan

Health Quality Council, Novartis, Nycomed, Pfizer; Employment: University of Saskatchewan, Saskatoon Health Region; Grants/grants pending (money to institution): Canadian Institute of Health Research, AstraZeneca, GlaxoSmithKline, Lung Association of Saskatchewan, Nycomed, Pfizer, Novartis, Saskatchewan Health Research Foundation, Schering-Plough, Saskatchewan Ministry of Health; Payment for lectures including service on speakers bureaus: AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Pfizer, Lung Association of Saskatchewan, Canadian Thoracic Society, American Thoracic Society. Dr. Wedzicha: Grants/grants pending (money to institution): Boehringer Ingelheim; Board membership: GlaxoSmithKline, Novartis, Bayer, Pfizer, Medimmune/Astra-Zeneca, Danone/Nutricia, Nycomed; Consultancy: Chiesi; Consultancy (money to institution): Novartis; Grants/grants pending (money to institution): GlaxoSmithKline, Novartis, Chiesi, AstraZeneca, Johnson & Johnson; Payment for lectures including service on speakers bureaus: Boehringer Ingelheim, GlaxoSmithKline, Pfizer, Bayer, Nycomed, Chiesi; Travel/accommodations/meeting expenses unrelated to activities listed: Boehringer Ingelheim. Dr. Shekelle: Employment: Veterans Affairs Medical Center; Grants/grants pending (money to institution): Agency for Healthcare Research and Quality, National Institutes of Health, Veterans Administration; Royalties: UpToDate; Travel/accommodations/meetings expenses unrelated to activities listed: Travel to meetings sponsored by AHRQ, the Health Foundation, the University of Michigan, VA, Italian regional health authority, and RAND. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M11-0925.

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

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: Moderate

1c.14 Summary of Controversy/Contradictory Evidence: No known areas of controversy.

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

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

Recommendation 1: ACP, ACCP, ATS, and ERS recommend that spirometry should be obtained to diagnose airflow obstruction in patients with respiratory symptoms (Grade: strong recommendation, moderate-quality evidence). Spirometry should not be used to screen for airflow obstruction in individuals without respiratory symptoms (Grade: strong recommendation, moderate-quality evidence). (Qaseem et al, 2011)

A clinical diagnosis of COPD should be considered in any patient who has dyspnea, chronic cough or sputum production, and/or a history of exposure to risk factors for the disease. Spirometry is required to make the diagnosis in this clinical context; the presence of a post-bronchodilator FEV1/FVC < 0.70 confirms the presence of persistent airflow limitation and thus of COPD. Whereas spirometry was previously used to support a diagnosis of COPD, spirometry is now required to make a confident diagnosis of COPD. Spirometry is the most reproducible and objective measurement of airflow limitation available. (GOLD 2015)

1c.17 Clinical Practice Guideline Citation: Amir Qaseem, MD, PhD, MHA; Timothy J. Wilt, MD, MPH; Steven E. Weinberger, MD; Nicola A. Hanania, MD, MS; Gerard Criner, MD; Thys van der Molen, PhD; Darcy D. Marciniuk, MD; Tom Denberg, MD, PhD; Holger Schunemann, MD, PhD, MSc; Wisia Wedzicha, PhD; Roderick MacDonald, MS; and Paul Shekelle, MD, PhD, for the American College of Physicians, the American College of Chest Physicians, the American Thoracic Society, and the European Respiratory Society. Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Ann Intern Med. 2011;155:179-191.

Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015. Available at http://www.goldcopd.org.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/content.aspx?id=34205&search=copd

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: See 1.c.10

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

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: Strong

1c.24 **Rationale for Using this Guideline Over Others:** It is the ATS policy to use guidelines which are evidence-based, applicable to physicians and other health-care providers, and developed by a national specialty organization or government agency.

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: Moderate **1c.26** Quality: Moderate **1c.27** Consistency: Moderate

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0091_Evidence_2015.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Despite major efforts to broadly disseminate the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines and use of COPD performance measures across different specialty societies, COPD remains underdiagnosed and misdiagnosed (Collins et al., 2015; Perez et al., 2011). Although spirometry use has increased, it remains underutilized to confirm airflow obstruction and accurately diagnose COPD (CDC, 2012; Nishi et al., 2013). Studies show proper COPD diagnosis with spirometry is done on just over half of patients in the US and Canada (Boulet et al., 2013; Bourbeau et al., 2008; Collins et al., 2015; Nishi et al., 2013; Perez et al., 2011; Yu et al., 2013) and ranges from 10-48% in the Asia-Pacific region, Africa, eastern Europe, and Latin America (Aisanov et al., 2012). A study of physician-diagnosed COPD patients hospitalized for exacerbations found that 22% of patients did not have COPD upon spirometry testing (Prieto Centurion, et al., 2012).

Treatment of COPD without accurate diagnosis and understanding of true etiology of symptoms results in patients not receiving medication that would improve symptoms and quality of life, prevent exacerbations and reduce costly use of emergency and hospital services while other patients may be exposed to adverse effects of unneeded medication and or delays in true diagnosis and management of another condition increasing overall cost of care (Boulet et al., 2013; Bourbeau et al., 2008; CDC, 2012; Collins et al., 2015; Joo et al., 2011). We believe this measure will continue to increase appropriate spirometry use to assist physicians in the accurate diagnosis and treatment of patients with COPD, improving patient management and reducing total costs of COPD.

Citations:

Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.

Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.

Bourbeau J, Sebaldt RJ, Day A, Bouchard J, Kaplan A, Hernandez P, Rouleau M, et al. Practice patterns in the management of chronic

obstructive pulmonary disease in primary practice: the CAGE study. Can Respir J. 2008 Jan-Feb:15(1):13-9.

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.

Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.

Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.

Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.

Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.

Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

This measure has been in use by the CMS Physician Quality Reporting Initiative/System (PQRI/S) since 2007 with the following reporting options:

- 2007 – Claims option
- 2008-2010, 2012, 2013 – Claims and registry options
- 2011 – Claims, registry and GPRO II options

Data from CMS(1) indicates a gap in care, trending favorably over time. Most recent data indicate a greater than 30% gap in care for 2014. This gap is aligned with research findings cited in 1b.3.

Average performance rate:
2010 - 56.0%
2011 - 68.3%
2012 - 69.4%
2013 - 53.4%
2014 - 67.1%

(1)Source: Timothy Jackson, CMS.

_____

Performance scores from 2012 comprehensive review submitted by PCPI to provide history.

CMS Physician Quality Reporting Initiative/System:
This measure was used in the CMS Physician Quality Reporting Initiative/System (PQRI/S) in the 2007 through 2011 claims option; 2009 through 2011 registry option; and the 2011 group practice reporting II option.

There is a gap in care as shown by this 2008 data; 45.7% of patients reported on did not meet the measure.(1)

10th percentile: 4.17%

25th percentile: 17.39%
50th percentile: 51.45%
75th percentile: 83.33%
90th percentile: 94.85%

Exception rate: 2.5%

(1) Confidential CMS PQRI Performance Information by Measure. Jan-Sept TAP file.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Studies show proper COPD diagnosis with spirometry is done on just over half of patients in the US and Canada (Boulet et al., 2013; Collins et al., 2015; Nishi et al., 2013; Perez et al., 2011; Yu et al., 2013,) and globally ranges from 6.5% in China to 59% in Sweeden with a mean of 26% in the Asia-Pacific region, Africa, eastern Europe, and Latin America (Aisanov et al., 2012; Yu et al., 2013).

Citations:

Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.

Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.

Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.

Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.

Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.

Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
We are not aware of any disparities data from this measure as specified. Please see 1b.5 for a summary of our findings in the literature regarding disparities.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Studies have been done to show associations between education level and income and outcomes related to COPD (Eisner et al., 2011; Holt et al., 2011). Studies also show association between gender and race on the incidence/severity of COPD (Bruse et al., 2011; Diaz et al., 2014; Foreman et al., 2011; Han et al., 2011). However, few research studies have been conducted to show disparities in use of spirometry.

One study showed misdiagnosis of COPD in an underserved, uninsured population. In a study of COPD patients from February 2011 to June 2012 at a federally qualified health center "eighty patients treated for a previous diagnosis of COPD (n = 72) or on anticholinergic inhalers (n = 8) with no COPD diagnosis were evaluated. The average age was 52.9 years; 71% were uninsured. Only 17.5% (14/80) of patients reported previous spirometry. Spirometry revealed that 42.5% had no obstruction, 22.5% had reversible obstruction, and 35% had nonreversible obstruction." Thus 42% of the patients were being over/mistreated (Ghattas et al., 2013).

Another study conducted in an outpatient primary clinic of a large urban hospital found no difference in use of spirometry between Causasians and minorities, or between normal weight and obese patients (Joo et al., 2011).

A review of COPD in Hispanics noted that common reasons for misdiagnosis in Hispanics may include lack of access to health care (which may include spirometry) and a high proportion of uninsured individuals (Brehm and Celedón, 2008).

The ATS is aware of health disparities related to respiratory diseases and has recently created a Health Equality Subcommittee of the Health Policy Committee. This group has been tasked with providing recommendations for moving toward respiratory health equality to include improving environmental factors, healthy lifestyle promotion, high quality healthcare (prevention, screening, diagnosis and treatment) and further research (Celedón et al., 2014).

Citations:

Brehm JM, Celedón JC. Chronic obstructive pulmonary disease in Hispanics. Am J Respir Crit Care Med. 2008 Mar 1;177(5):473-8.

Bruse S, Sood A, Petersen H, Liu Y, Leng S, Celedón JC, Gilliland F, Celli B, Belinsky SA, Tesfaigzi Y. New Mexican Hispanic smokers have lower odds of chronic obstructive pulmonary disease and less decline in lung function than non-Hispanic whites. Am J Respir Crit Care Med. 2011 Dec 1;184(11):1254-60.

Celedón JC, Roman J, Schraufnagel DE, Thomas A, Samet J. Respiratory health equality in the United States. The American thoracic society perspective. Ann Am Thorac Soc. 2014 May;11(4):473-9.

Diaz AA, Come CE, Mannino DM, Pinto-Plata V, Divo MJ, Bigelow C, Celli B, Washko GR. Obstructive lung disease in Mexican Americans and non-Hispanic whites: an analysis of diagnosis and survival in the National Health and Nutritional Examination Survey III Follow-up Study. Chest. 2014 Feb;145(2):282-9.

Eisner MD, Blanc PD, Omachi TA, Yelin EH, Sidney S, Katz PP, Ackerson LM, Sanchez G, Tolstykh I, Iribarren C. Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health 2011;65:26–34.

Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, et al. Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene Study. Am J Respir Crit Care Med. 2011 Aug 15;184(4):414-20.

Ghattas C, Dai A, Gemmel DJ, Awad MH. Over diagnosis of chronic obstructive pulmonary disease in an underserved patient population. Int J Chron Obstruct Pulmon Dis. 2013;8:545-9.

Han MK, Curran-Everett D, Dransfield MT, Criner GJ, Zhang L, Murphy JR, Hansel NN, DeMeo DL, Hanania NA, Regan EA, Make BJ, Martinez FJ, Westney GE, Foreman MG; COPDGene Investigators. Racial differences in quality of life in patients with COPD. Chest. 2011 Nov;140(5):1169-76.

Holt JB, Zhang X, Presley-Cantrell L, Croft JB. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. Int J Chron Obstruct Pulmon Dis. 2011; 6 321–328.

Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

A 2013 analysis estimates the COPD prevalence in US adults aged 40-79 to be 14-15% (Tiler et al., 2013). Significant mortality and morbidity and healthcare costs are associated with COPD.

COPD is the third ranked cause of death in the US and the number of deaths due to COPD increased 3.5% between 2010 and 2011 (CDC, 2012; Heron, 2015).

Worldwide, the Global Burden of Disease estimates 328 million people with COPD including 168 million men and 160 million women. COPD was the sixth ranked cause of death in 1990, moving upward to fourth in 2000 and estimated to be the third leading cause of death globally by 2020 (Lopez-Campos et al., 2015). The World Health Organization estimates 3 million people with COPD in the world die each year due to COPD (Diaz-Guzman and Mannino, 2014). COPD exacerbations resulting in hospitalizations correlate with 1-year mortality of 21% and 5-year mortality of 55% (Lopez-Campos et al., 2015).

Disability as a result of COPD is high globally with 29.4 million [Years Lost to Disability] YLD. YLD due to COPD rankings have moved from sixth to fifth from 1990 to 2010 (Lopez-Campos et al., 2015).

COPD exacerbations impact quality of life and healthcare costs. The cost of hospitalizations related to COPD exacerbations in the US is estimated to be $18 billion annually (Lopez-Campos et al., 2015). Thirty day readmission rates for patients discharged from hospitals for COPD exacerbations are 20% and 35% at 90-days (Prieto Centurion et al., 2012).

Mathematica Policy Research conducted an analysis on CMS data in 2012 finding that per capita costs for beneficiaries with COPD was $24,901, compared to overall per capita costs of $10,734 (Mathematica Policy Research, 2014).

Despite the high prevalence, mortality, morbidity and cost associated with COPD, underdiagnosis remains a key challenge worldwide (Lopez-Campos et al., 2015).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Citations:

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clin Chest Med. 2014 Mar 35(1):7-16.

Heron M. Deaths: leading causes for 2011. Natl Vit Stat Rep. 2015 Jul 27;64(7):1-96

Lopez-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2015 Oct 23. [Epub ahead of print].

Mathematica Policy Research. Experience Report for the Performance Year 2012 Quality and Resource Use Reports. January 8, 2014. Accessed December 7, 2015. Accessible at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2012-QRUR_Experience_Report.pdf

Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.

Tiler T, Dillon C, Paulose-Ram R, Hnizdo E, Doney B. Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post- bronchodilator spirometry: the National Health and Nutrition Examination Survey (NHANES) 2007-2010. Respir Res. 2013 Oct 9;14:103.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input*

was obtained.)
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
The updated specifications for this measure are included within this form.

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
None

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients with documented spirometry results in the medical record (FEV1 and FEV1/FVC)

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once per reporting period

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Numerator Quality-Data Coding Options for Reporting Satisfactorily

Numerator Instructions: Look for most recent documentation of spirometry evaluation results in the medical record; do not limit the search to the reporting period.

To submit the numerator option for spirometry results documented and reviewed, report the following:

Performance Met: CPT II 3023F: Spirometry results documented and reviewed

OR

Spirometry Results not Documented for Medical, Patient, or System Reasons
Append a modifier (1P, 2P or 3P) to CPT Category II code 3023F to report documented circumstances that appropriately exclude patients from the denominator.

Medical Performance Exclusion: 3023F with 1P: Documentation of medical reason(s) for not documenting and reviewing spirometry results

OR

Patient Performance Exclusion: 3023F with 2P: Documentation of patient reason(s) for not documenting and reviewing spirometry results

OR

System Performance Exclusion: 3023F with 3P: Documentation of system reason(s) for not documenting and reviewing spirometry results

OR

Spirometry Results not Documented, Reason not Otherwise Specified
Append a reporting modifier (8P) to CPT Category II code 3023F to report circumstances when the action described in the numerator is not performed and the reason is not otherwise specified.

Performance Not Met: 3023F with 8P: Spirometry results not documented and reviewed, reason not otherwise specified

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
All patients aged 18 years and older with a diagnosis of COPD

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
All Patients aged >= 18 years on date of encounter

AND
Diagnosis for COPD (ICD-10-CM): J41.0, J41.1, J41.8, J42, J43.0, J43.1, J43.2, J43.8, J43.9, J44.0, J44.1, J44.9

AND

Patient encounter during the reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Documentation of medical reason(s) for not documenting and reviewing spirometry results
Documentation of patient reason(s) for not documenting and reviewing spirometry results
Documentation of system reason(s) for not documenting and reviewing spirometry results

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1*

page should be provided in an Excel or csv file in required format at S.2b)
ATS continues to use the PCPI exception methodology that uses three categories of exception reasons for which a patient may be removed from the denominator of an individual measure: medical, patient and system reasons.

Exceptions are used to remove patients from the denominator of a performance measure when a patient does not receive a therapy or service AND that therapy or service would not be appropriate due to specific reasons; otherwise, the patient would meet the denominator criteria. Exceptions are not absolute, and the application of exceptions is based on clinical judgment, individual patient characteristics, or patient preferences. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions include medical reason(s), patient reason(s) or system reason(s) for not documenting spirometry results. Although this methodology does not require the external reporting of more detailed exception data, the ATS recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The ATS also conducts systematic review and analysis of exceptions data to identify practice patterns and opportunities for quality improvement.

For Claims:
Documentation of medical, patient, or system reason(s) for not documenting and reviewing spirometry results.

Append a modifier (1P, 2P or 3P) to CPT Category II code 3023F to report documented circumstances that appropriately exclude patients from the denominator.
3023F with 1P: Documentation of medical reason(s) for not documenting and reviewing spirometry results
3023F with 2P: Documentation of patient reason(s) for not documenting and reviewing spirometry results
3023F with 3P: Documentation of system reason(s) for not documenting and reviewing spirometry results

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
We encourage the results of this measure to be stratified by race, ethnicity, primary language, and administrative sex.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
No risk adjustment or risk stratification.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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.)

1. Start with Denominator

2. Check Patient Age:

a. If the Age is greater than or equal to 18 years of age on Date of Service and equals No during the measurement period, do not include in Eligible Patient Population. Stop Processing.

b. If the Age is greater than or equal to 18 years of age on Date of Service and equals Yes during the measurement period, proceed to check Patient Diagnosis.

3. Check Patient Diagnosis:

a. If Diagnosis of COPD as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.

b. If Diagnosis of COPD as Listed in the Denominator equals Yes, proceed to check Encounter Performed.

4. Check Encounter Performed:

a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.

b. If Encounter as Listed in the Denominator equals Yes, include in the Eligible population.

5. Denominator Population:

a. Denominator population is all Eligible Patients in the denominator. Denominator is represented as Denominator in the Sample Calculation listed at the end of this document. Letter d equals 8 patients in the sample calculation.

6. Start Numerator

7. Check Spirometry Results Documented and Reviewed:

a. If Spirometry Results Documented and Reviewed equals Yes, include in Reporting Met and Performance Met.

b. Reporting Met and Performance Met letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter a equals 4 patients in Sample Calculation.

c. If Spirometry Results Documented and Reviewed equals No, proceed to Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results.

8. Check Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results:

a. If Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.

b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b1 equals 1 patient in the Sample Calculation.

c. If Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results.

9. Check Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results:

a. If Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.

b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b2 equals 0 patients in the Sample Calculation.

c. If Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results.

10. Check Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results:

a. If Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.

b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b3 equals 0 patients in the Sample Calculation.

c. If Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Spirometry Results Not Documented and Reviewed, Reason Not Specified.

11. Check Spirometry Results Not Documented and Reviewed, Reason Not Specified:

a. If Spirometry Results Not Documented and Reviewed, Reason Not Specified equals Yes, include in Reporting Met and

Performance Not Met.
b. Reporting Met and Performance Not Met letter is represented in the Reporting Met in the Sample Calculation listed at the end of document. Letter c equals 2 patients in the Sample Calculation.
c. If Spirometry Results Not Documented and Reviewed, Reason Not Specified equals No, include in Reporting Not Met.

12. Check Reporting Not Met
a. If Reporting Not Met equals No, Quality Data Code or equivalent not reported. 1 patient has been subtracted from the reporting numerator in sample calculation.

Please see Measure Flow in Appendix A.1 for ´Sample Calculation´ referenced above.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure does not require sampling or a survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not Applicable

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Team

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
This is not a composite measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0091_MeasureTesting_MSF5.0_Data.doc,0091_testing_attachment_2015_amended_122915.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0091
Measure Title: COPD: Spirometry Evaluation
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion

impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

NOTE: As requested, test responses from the 2012 comprehensive review testing form submitted by the AMA-PCPI are included in black font. Due to differences in form structure, sections and questions, and due to different testing, verbatim responses are copied where determined to be most appropriate. What appear to be errors in alignment of responses to questions and/or omissions are likely and are due to these differences. If necessary for clarification, please refer to original 2012 testing form.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Refer to the validity section for a description of the data sample for our EHR testing project.

EHR Measure Validity
The measure was calculated using data collected using two different methods of collection:
- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

The data source was electronic health records in an ambulatory care setting.
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 123 patient encounters.
Data collected from patients seen between 01/01/2010-12/31/2011.
Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

Face Validity
An expert panel was used to assess face validity of the measure. This panel consisted of 12 members, with representation from a number of specialties, including internal medicine, methodology, pulmonology, family medicine, critical care medicine, emergency medicine, pharmacy science, nursing, and health plan representation.

Co-Chairs:
William E. Golden, MD, FACP (University of Arkansas College of Medicine)
Linus Santo Tomas, MD, MS (American College of Chest Physicians)

Members:
Bruce Bagley, MD (American Academy of Family Physicians)
Troy T. Fiesinger, MD (American Academy of Family Physicians)
David G. Jaimovich, MD (Society of Critical Care Medicine)
Bruce Krieger, MD (American Thoracic Society)
Thomas W. Lukens, MD, PhD, FACEP (American College of Emergency Physicians)
Deborah Patterson, MS, RN (Blue Cross Blue Shield Association)
Sam J. W. Romeo, MD, MBA (Tower Health & Wellness Center)
Ralph M. Schapira, MD (VA Medical Center)
Sean D. Sullivan, RPh, PhD (Department of Pharmacy, University of Washington)
Dennis E. Richling, MD (Midwest Business Group on Health)

2015 submission

The data source is the Centers for Medicare & Medicaid Services (CMS) Medicare administrative claims database.

The testing was conducted by Mathematica Policy Research as a component of the 2012 Quality and Resource Use Report (QRUQ), part of the CMS Physician Feedback Reporting Program.

Citation:

Mathematica Policy Research. Experience Report for the Performance Year 2012 Quality and Resource Use Reports. January 8, 2014. Accessed December 7, 2015. Accessible at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2012-QRUR_Experience_Report.pdf

Note that this measure was also tested in 2012 by the PCPI to support NQF re-endorsement for the 2012 comprehensive review. Those test results are not repeated here, however, are available as an attachment on the NQF submission form.

1.3. What are the dates of the data used in testing? January 2012 – December 2012

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**

☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Testing and analysis included 2,064 groups of physicians with at least 25 eligible professionals (EPs) (average of 120 EPs per group). Of these, there were 693 groups of physicians with at least 100 EPs (average of 322 EPs). This group represents 30% of medical group practices with 25 or more EPs nationwide. Groups were included if they reported at least 20 eligible cases for the measure. The groups were distributed across all states, the District of Columbia, Guam and Puerto Rico.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Testing and analysis included 11,593,241 Medicare beneficiaries identified on claims associated with the groups described in 1.5. Beneficiaries attributed to groups with more than 25 EPs averaged 2,974 (standard deviation = 5,105). Approximately half (52%) of the groups were attributed fewer than 1,000 beneficiaries. Beneficiaries attributed to groups with more than 100 EPs averaged 7,077 (standard deviation = 7,842).

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The data were used for reliability testing only. Face validity testing was done with a survey. Other analyses were not done or not applicable.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patients in the testing and analysis were Medicare beneficiaries. No other sociodemographic variables were available for analysis.

2a2. RELIABILITY TESTING

Note: *If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Refer to the validity section for a description of the analytic methods for our EHR testing project.

2015 submission

The method of reliability testing as used by Mathematica Policy Research is described as:

"For each of these measures, reliability was estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error):

"Reliability = Variation between groups/(Variation between groups + Variation within group)

"If a score is deemed highly reliable, we would expect that a group's performance rates would be very similar if performance were calculated on the basis of a random sample of the practice's beneficiaries.

"Reliability scores are represented on a continuum from zero to one. Scores closer to zero indicate lower reliability and scores closer to one indicate higher reliability. Although there is no universally agreed-upon minimum reliability threshold, reliability scores in the 0.40–0.70 range are often considered moderate, and scores greater than 0.70 are considered high."

Please see 1.2 for citation.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Refer to the validity section for the testing results for our EHR testing project.

2015 submission

As noted above, scores above 0.70 are considered high.
The reliability for this measure among groups with 25 or more EPs was 0.73.
The reliability for this measure among groups with 100 or more EPs was 0.83.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

We believe this measure remains reliable based on high reliability test scores and relatively large test sample size. We also believe that the measure is reliable across relatively small groups and relatively large groups.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:
• Percent agreement at the denominator and numerator(exception - for those measures with exception)
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

Face Validity
Face validity of the measure score as an indicator of quality was systematically assessed as follows.
After the measure was fully specified, the expert panel (workgroup membership) was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Scale 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5= Strongly Agree

2015 submission

Face validity of the measure score as an indicator of quality was systematically assessed using the following approach:

After the measure was fully specified, the ATS Clinical Practice Committee was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

The rating scale used was 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

The 12 members of the ATS COPD Clinical Practice Committee were selected to serve as an expert panel:

Kevin L. Kovitz,, MD
Robert DeMarco, MD
Scott Manaker, MD
Michael Donahoe, MD
Omar Hussain, MD
Katina Nicolacakis, MD
Tom Gildea, MD
Steve G. Peters, MD
Kashif Hussain, MD
Stephen Hoffman, MD
Alan Plummer, MD
Mike Nelson, MD

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
This measure demonstrates substantial agreement when comparing the automated EHR report to visual inspection.

Reliability: N, % Agreement, Kappa
Numerator: 123, 86.89%, 0.7281 (0.6086-0.8476 CI)
Denominator: 123, 100%, kappa non-calculable (non-calculable CI)*

*Kappa statistic could not be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

Face Validity
The results of the expert panel rating of the validity statement were as follows: N = 7; Mean rating = 4.86 and 100% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings
1 - 0 (Strongly Disagree)
2 - 0
3 - 0 (Neither Agree nor Disagree)
4 - 1
5 - 6 (Strongly Agree)

2015 submission

The results of the expert panel rating of the validity statement include:

- N = 12
- Mean rating = 4.6
- Panelists that agree or strongly agree that this measure can accurately distinguish good and poor quality = 91.7%

Frequency distribution of ratings

1 - Strongly disagree	0
2	0
3 - Neither Agree nor Disagree	1
4	3
5 - Strongly Agree	8

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

We believe this measure remains valid based on the degree of agreement by a panel of testers.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 123 patient encounters.
Data collected from patients seen between 01/01/2010-12/31/2011.
Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

Exceptions included medical, patient and system reasons. Exceptions were analyzed for frequency and variability across providers.

2015 submission

Exclusion analysis was not conducted on this measure in this study.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
Exception rate: 0.81%
Validity of exceptions was 0% agreement with a kappa of 0.0000*

*Due to the small sample size and the single exception found during manual abstraction, the resulting agreement rate and kappa statistic are low.

2015 submission

Not available

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not available

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

CMS Physician Quality Reporting Initiative/System:
98,074 cases were reported on for the 2008 program, the most recent year for which data is available.

The following information is for the 2009 program, the only year for which such data is available.
Clinical Condition and Measure: #51 Spirometry Evaluation
Eligible Professionals: 212,885
Professionals Reporting: 1,841
% Professionals Reporting: 0.86%
# Professionals Reporting >=80% of eligible instances: 737
% Professionals Reporting >=80% of eligible instances: 40.03%

CMS Physician Quality Reporting Initiative/System:
The inter-quartile range (IQR) was calculated to determine the variability of performance on the measure.

2015 submission

Analysis of the differences in performance rates was conducted through benchmarks. According to Mathematica Policy Research, "Prior-year benchmarks were also computed for the claims-based quality indicators, and none of the measures differed significantly at the 5 percent level from the prior year benchmark. A weighted average (based on eligible cases) of performance for groups with 25 or more EPs serves as the benchmark for all groups of this size, whereas a comparable weighted average among groups with at least 100 EPs forms the benchmark for larger groups (100 or more EPs)."

Please see 1.2 for citation.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

CMS Physician Quality Reporting Initiative/System:
Scores on this measure: N = 98,074; Mean = 54.30%,
10th percentile: 4.17%
25th percentile: 17.39%
50th percentile: 51.45%
75th percentile: 83.33%
90th percentile: 94.85%

The inter-quartile range (IQR) provides a measure of the dispersion of performance. The IQR is 65.94 and indicates that 50% of physicians have performance on this measure ranging from 17.39% and 83.33% and 10% of physicians have performance rates less than or equal to 4.17%.(1)

(1)Confidential CMS PQRI 2008 Performance Information by Measure. Jan-Sept TAP file.

2015 submission

The percent of groups different than the benchmark (p<0,05) for this measure among groups with 25 or more EPs was 45.6%.
The percent of groups different than the benchmark (p<0,05) for this measure among groups with 100 or more EPs was 47.1%.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The proportion of groups statistically different than the benchmark suggests that there is variation across group performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

The measure was calculated using data collected using two different methods of collection:
- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:
• Percent agreement at the denominator and numerator(exception - for those measures with exception)
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

2015 submission

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
This measure demonstrates substantial agreement when comparing the automated EHR report to visual inspection.

Reliability: N, % Agreement, Kappa
Numerator: 123, 86.89%, 0.7281 (0.6086-0.8476 CI)
Denominator: 123, 100%, kappa non-calculable (non-calculable CI)*

*Kappa statistic could not be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

2015 submission

Not applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Missing data analysis was not conducted on this measure in this study.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Not available

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not available

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

This measure has been in use in the PQRI/S program since 2007. We continue to find this measure reliable and feasible for implementation.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

N/A

RESPONSE TO ITEM 3b.2

ATS is considering eMeasures in the future. At this time we have not determined whether this measure will be converted.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the

time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Quality Improvement (Internal to the specific organization)	Public Reporting Physician Compare https://www.medicare.gov/physiciancompare/staticpages/data/aboutthedata.html Payment Program CMS PQRS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

This measure is planned for integration into the CMS Physician Compare Program. Although Physician Compare has been launched, this measure is not yet included as of 12/14/15. The purpose of the Physician Compare Program is to help Medicare beneficiaries make informed choices about health care. The program is broadly available through the Physician Compare website noted above.

This measure has been in use for the CMS PQRS program since 2007. The PQRS is a quality reporting program to encourage individual eligible professionals and group practices to report quality information to Medicare. Effective in 2015, the PQRS will be used to apply a negative payment adjustment to individual eligible professionals and group practices who do not satisfactorily report data on quality measures for covered professional services provided to Medicare patients in 2013.

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

N/A

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Please see 1b.2 for performance results, including favorable trends from 2011-2014. Please see 1b.5 for research results on geographical variance in performance. Please see 1b.4 for performance on disparities. We are aware of no further analyses conducted on improvement.

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of

initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
This measure was initially developed for use in the PQRI/S program. Although not publically reported until the recent release of the new Physician Compare site, the performance trends indicate progress on increasing use of spirometry to accurately diagnose COPD from 2011-2014.

The ATS supports the goal of high-quality, efficient healthcare. Toward that goal, the ATS Quality Improvement Committee reviews performance annually as a component of measure maintenance and plans further analyses in the future. ATS participates in international COPD guideline development, publishes technical standards for spirometry and periodically develops tools and educational material to support members in quality improvement. ATS also conducts educational sessions and an annual meeting featuring use of guidelines, including appropriate spirometry use.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We are not aware of any unintended consequences related to this measure.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0577 : Use of Spirometry Testing in the Assessment and Diagnosis of COPD

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
These measures have distinct differences in their denominators and numerators. First, our measure is broader in denominator population, being for all patients age 18 years and older with a diagnosis of COPD, while 0577 is for patients age 40 years and older with a new diagnosis of COPD. Our measure is more consistent with COPD guidelines, which do not state an age to start using a spirometry evaluation; rather, spirometry should be used to assess all adults with COPD, not just adults with a new diagnosis of

COPD. Second, our measure´s numerator is more flexible than 0577, allowing a spirometry evaluation anytime during the measurement period, rather than 0577´s requirement that spirometry be performed within 6 months of a new diagnosis of COPD. Our measure numerator is also specific to spirometry results, requiring both the FEV1/FVC values.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 0091_measure_flow_2015_form_no_s.19.doc

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Thoracic Society
Co.2 **Point of Contact:** Gary, Ewart, gewart@thoracic.org, 202-296-9770-
Co.3 **Measure Developer if different from Measure Steward:** Northfield Associates LLC
Co.4 **Point of Contact:** Sue, Frechette, sue.frechette@northfieldassoc.com, 802-496-7815-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
This measure was initially developed in 2007 by the AMA-PCPI, working with the ATS.

PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

The initial Work Group Panel consisted of:
William E. Golden, MD, FACP, co-chair
Linus Santo Tomas, MD, MS, co-chair
Bruce Bagley, MD (AAFP)
Troy T. Fiesinger, MD (AAFP)
David G. Jaimovich, MD (SCCM)
Bruce Krieger, MD (ATS)
Thomas W. Lukens, MD, PhD, FACEP (ACEP)
Susan Nedza, MD, MBA, FACEP (CMS)
Deborah Patterson, MS, RN (BCBSA)
Sam J. W. Romeo, MD, MBA
Ralph M Schapira, MD (VA)
Sean D. Sullivan, RPh, PhD

Dennis E. Richling, MD
Nancy Lawler, RN (Joint Commission)

Stewardship of this measure was transferred to the ATS in November 2014.

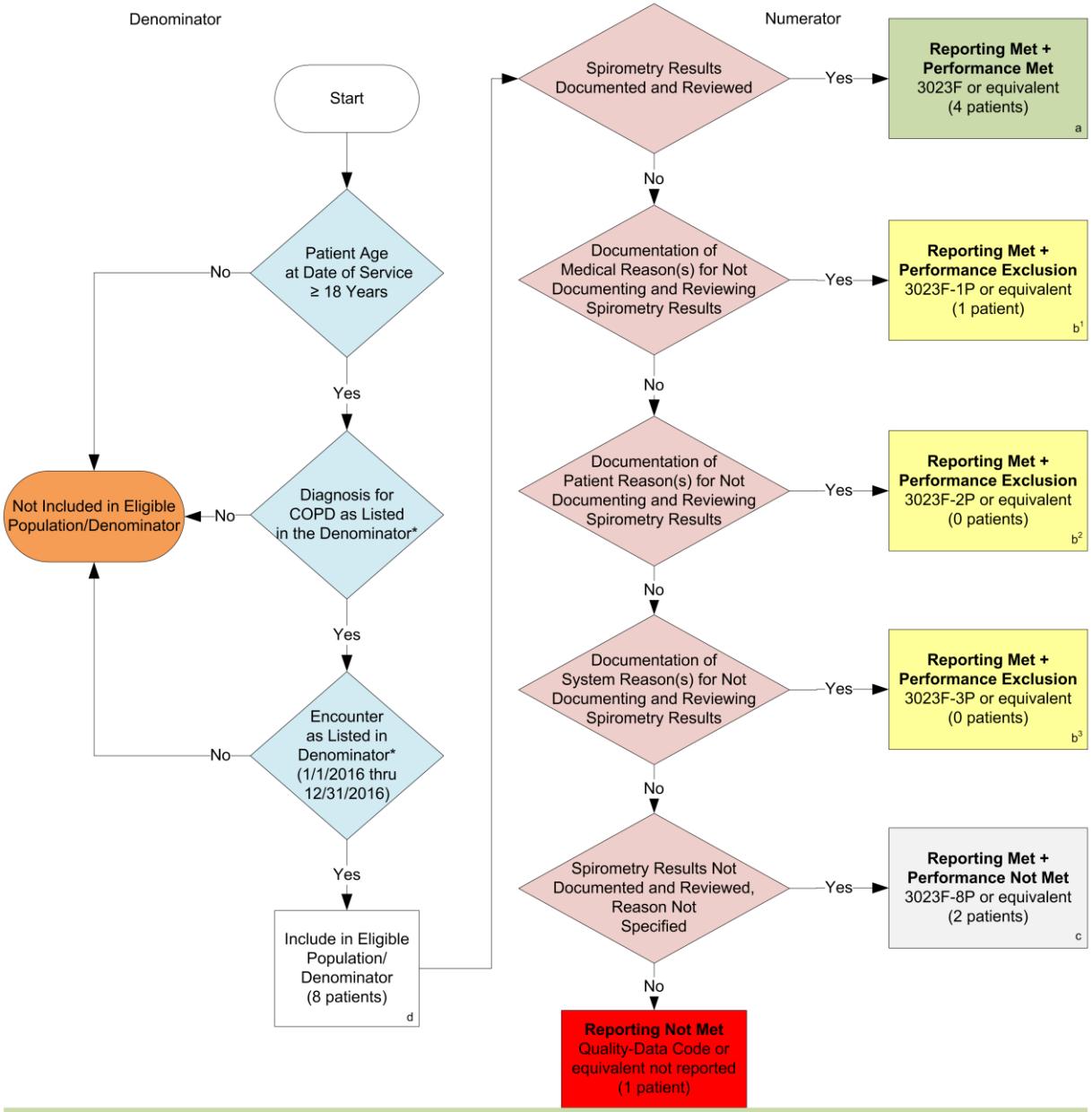
To prepare for the 2015 NQF comprehensive review, ATS formed the Quality Improvement Committee Sub-committee on COPD Measures to review and update this measure. The Sub-committee members include:
Laura Feemster, MD, MS, VA Puget Sound Health Care System, University of Washington Medical Center, Chair
Bela Patel, MD, The University of Texas Health Science Center at Houston
Carolyn Fruci, MD, PhD, Prima-CARE, PC
David Au, MD, MS, VA Puget Sound Health Care System, University of Washington Medical Center
Jerry A. Krishnan, MD, PhD, University of Illinois at Chicago
Gary Ewart, Chief, Advocacy & Government Relations, ATS
Sue Frechette, RN, MBA, Consultant, Northfield Associates LLC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2007
Ad.3 **Month and Year of most recent revision:** 12, 2015
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** Coding/Specifications updates occur annually. ATS plans to continue measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. ATS will also review the measures if there is a major change in scientific evidence, results from testing or other issues are noted that materially affect the integrity of the measure.

2016 Claims/Registry Individual Measure Flow
PQRS #51 NQF #0091: Chronic Obstructive Pulmonary Disease (COPD): Spirometry Evaluation



SAMPLE CALCULATIONS:

Reporting Rate=

Performance Met (a=4 patients) + Performance Exclusion ($b^1+b^2+b^3$=1 patient) + Performance Not Met (c=2 patients) / Eligible Population / Denominator (d=8 patients) = 7 patients / 8 patients = **87.50%**

Performance Rate=

Performance Met (a=4 patients) / Reporting Numerator (7 patients) – Performance Exclusion ($b^1+b^2+b^3$=1 patient) = 4 patients / 6 patients = **66.67%**

*See the posted Measure Specification for specific coding and instructions to report this measure.
NOTE: Reporting Frequency – Patient-intermediate

v2

2016 Claims/Registry Individual Measure Flow
PQRS #51 NQF #0091: Chronic Obstructive Pulmonary Disease (COPD): Spirometry Evaluation

Please refer to the specific section of the Measure Specification to identify the denominator and numerator information for use in reporting this Individual Measure.

1. Start with Denominator
2. Check Patient Age:
    a. If the Age is greater than or equal to 18 years of age on Date of Service and equals No during the measurement period, do not include in Eligible Patient Population. Stop Processing.
    b. If the Age is greater than or equal to 18 years of age on Date of Service and equals Yes during the measurement period, proceed to check Patient Diagnosis.
3. Check Patient Diagnosis:
    a. If Diagnosis of COPD as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Diagnosis of COPD as Listed in the Denominator equals Yes, proceed to check Encounter Performed.
4. Check Encounter Performed:
    a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Encounter as Listed in the Denominator equals Yes, include in the Eligible population.
5. Denominator Population:
    a. Denominator population is all Eligible Patients in the denominator. Denominator is represented as Denominator in the Sample Calculation listed at the end of this document. Letter d equals 8 patients in the sample calculation.
6. Start Numerator
7. Check Spirometry Results Documented and Reviewed:
    a. If Spirometry Results Documented and Reviewed equals Yes, include in Reporting Met and Performance Met.
    b. Reporting Met and Performance Met letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter a equals 4 patients in Sample Calculation.
    c. If Spirometry Results Documented and Reviewed equals No, proceed to Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results.
8. Check Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results:
    a. If Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.
    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b^1 equals 1 patient in the Sample Calculation.
    c. If Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results.
9. Check Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results:
    a. If Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.
    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b^2 equals 0 patients in the Sample Calculation.

c. If Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results.

10. Check Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results:

    a. If Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b^3 equals 0 patients in the Sample Calculation.

    c. If Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Spirometry Results Not Documented and Reviewed, Reason Not Specified.

11. Check Spirometry Results Not Documented and Reviewed, Reason Not Specified:

    a. If Spirometry Results Not Documented and Reviewed, Reason Not Specified equals Yes, include in Reporting Met and Performance Not Met.

    b. Reporting Met and Performance Not Met letter is represented in the Reporting Met in the Sample Calculation listed at the end of document. Letter c equals 2 patients in the Sample Calculation.

    c. If Spirometry Results Not Documented and Reviewed, Reason Not Specified equals No, include in Reporting Not Met.

12. Check Reporting Not Met

    a. If Reporting Not Met equals No, Quality Data Code or equivalent not reported. 1 patient has been subtracted from the reporting numerator in sample calculation.

SAMPLE CALCULATIONS:

Reporting Rate=

$$\frac{\text{Performance Met (a=4 patients) + Performance Exclusion } (b^1+b^2+b^3=1 \text{ patient}) + \text{Performance Not Met (c=2 patients)}}{\text{Eligible Population / Denominator (d=8 patients)}} = \frac{7 \text{ patients}}{8 \text{ patients}} = \mathbf{87.50\%}$$

Performance Rate=

$$\frac{\text{Performance Met (a=4 patients)}}{\text{Reporting Numerator (7 patients)} - \text{Performance Exclusion } (b^1+b^2+b^3=1 \text{ patient})} = \frac{4 \text{ patients}}{6 \text{ patients}} = \mathbf{66.67\%}$$

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0102
Measure Title: COPD: inhaled bronchodilator therapy
Measure Steward: American Thoracic Society
Brief Description of Measure: Percentage of patients aged 18 years or older, with a diagnosis of COPD (FEV1/FVC < 70%) who have an FEV1 < 60% predicted and have symptoms who were prescribed a long-acting inhaled bronchodilator
Developer Rationale: Despite major efforts to broadly disseminate the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines and use of COPD performance measures across different specialty societies, management of COPD, and specifically prescription for long-acting inhaled bronchodilators, remains suboptimal. Studies show a wide range of deficiencies in adherence to guidelines regarding long-acting inhaled bronchodilator use across different settings (Asche et al., 2012; CDC, 2012; Fitch, et al., 2011; Nantsupawat et al., 2012; Perez et al., 2011; Sharif, et al., 2013). Underuse of broncholidators were found related to hospital readmissions and to increased total costs of services when compared to patient care adhering to GOLD guidelines (Asche et al., 2012; Nantsupawat et al., 2012).

Suboptimal COPD management has implications for severity of illness, disease progression, patient quality of life and health status, exacerbations (and associated costs) and mortality. Improved adherence to COPD management guidelines, specifically appropriate use of long-acting inhaled bronchodilators, has the potential to improve clinical outcomes and cost of care related to COPD. As a result, we believe this measure will continue to increase appropriate long-acting inhaled bronchodilator use, improving patient management and total costs of COPD.

Citations:

Asche CV, Leader S, Plauschinat C, Raparla S, Yan M, Ye X, Young D. Adherence to current guidelines for chronic obstructive pulmonary disease (COPD) among patients treated with combination of long-acting bronchodilators or inhaled corticosteroids. Int J Chron Obstruct Pulmon Dis. 2012;7:201-9.

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Fitch K, Iwasaki K, Pvenson B. Plauschinat C, Zhang J. Variation in adherence with Global Initiative for Chronic Obstructive Lung Disease (GOLD) drug therapy guidelines: a retrospective actuarial claims data analysis. Curr Med Res Opin. 2011 Jul;27(7):1425-9.

Nantsupawat T, Limsuwat C, Nugent K. Factors affecting chronic obstructive pulmonary disease early rehospitalization. Chron Respir Dis. 2012 May;9(2):93-8.

Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.

Sharif R, Cuevas CR, Wang Y, Arora M, Sharma G. Guideline adherence in management of stable chronic obstructive pulmonary disease. Respir Med. 2013 Jul;107(7):1046-52.

Numerator Statement: Patients who were prescribed a long-acting inhaled bronchodilator
Denominator Statement: All patients aged 18 years and older with a diagnosis of COPD, who have FEV1/FVC < 70%, FEV1 <60% predicted and have symptoms (eg, dyspnea, cough/sputum, wheezing)
Denominator Exclusions: ATS continues to use the PCPI exception methodology that uses three categories of exception reasons for

which a patient may be removed from the denominator of an individual measure: medical, patient and system reasons.

Exceptions are used to remove patients from the denominator of a performance measure when a patient does not receive a therapy or service AND that therapy or service would not be appropriate due to specific reasons; otherwise, the patient would meet the denominator criteria. Exceptions are not absolute, and the application of exceptions is based on clinical judgment, individual patient characteristics, or patient preferences. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions include medical reason(s), patient reason(s) or system reason(s) for not prescribing long-acting inhaled bronchodilators. Although this methodology does not require the external reporting of more detailed exception data, the ATS recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice, Clinician : Team

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- During the previous review, the Committee agreed the evidence was appropriate and consistent for the use of spirometry to confirm the diagnosis of Chronic Obstructive Pulmonary Disease (COPD), however, it expressed concerns about the strength of evidence for the range of 60-70% FEV1/FVC ratio.
- Updated evidence for this process measure is based on 3 clinical practice guidelines for the diagnosis and management of Chronic Obstructive Lung Disease.
    - Dated 2015, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines provide ungraded recommendations for COPD assessment. The GOLD guidelines referenced 613 studies to update the previous set of guidelines from 2013.
    - The American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society 2011 guidelines graded the evidence for stable and symptomatic patients as strong recommendations with moderate quality evidence (the second highest ranking in this grading system). The recommendations are:
        - "For stable COPD patients with respiratory symptoms and FEV1 <60% predicted, ACP, ACCP, ATS, and ERS recommend treatment with inhaled bronchodilators."
        - "…recommend that clinicians prescribe monotherapy using either long-acting inhaled

anticholinergics or long-acting inhaled ß-agonists for symptomatic patients with COPD and FEV1 <60% predicted."

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates: Guidelines republished since last review. The 2015 GOLD guidelines referenced 613 studies to update the previous set of guidelines from 2013. The 2011 American College of Physicians et al. guidelines referenced 62 studies to update the previous set of guidelines from 2007.

Exception to evidence: Not applicable

Guidance from the Evidence Algorithm: 1→3 →4→5 (eligible for HIGH rating)

Questions for the Committee:

- *Although the guidelines have been updated, the underlying evidence presented appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reported this measure was used in the CMS Physician Quality Reporting Initiative/System (PQRS): 2007 through 2013 claims option; 2009 through 2013 registry option; 2011 through 2012 group practice reporting II option; and the 2012 ACO option. In the 2008 data, 53.61% of patients reported on did not meet the measure.
- The developer presented the following average performance rate:
  - 2010 - 89.7%
  - 2011 - 73.4%
  - 2012 - 98.5%
  - 2013 - 97.0%
  - 2014 - 95.9%

Disparities

- The developer reports disparities are identified as an issue in the literature, but results for this measure's ability to detect them were not provided.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Is this measure "topped out" and so should be considered for Reserve Status?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**Process measure based on claims, EHR. Level of analysis is clinician (group/team).
Developer reports evidence of systematic review, evaluation of quality, quantity and consistency of the evidence and

that the evidence is graded.
Updated evidence provided (2 clinical practice guidelines)--1) GOLD 2015 (613 studies to update prior version), 2) ACP/ACCP/ATS/ERS 2011 COPD mgt -- strong recommendation with moderate quality evidence (2nd highest ranking in this system). This has the strongest statement re: COPD with FEV1 < 60%. An additional 62 studies updated from the prior 2007 guideline.
Guidance from the Evidence Algorithm: 1→3 →4→5 (eligible for HIGH rating)
Questions for the Committee:
o Although the guidelines have been updated, the underlying evidence presented appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence? YES
**Process measure associated with improved treatment outcomes based on high quality evidence.
**I agree and so there is no need for repeat discussion and vote on Evidence.

1b. Performance Gap
Comments:
**• The developer reported this measure was used in the CMS Physician Quality Reporting Initiative/System (PQRS): 2007 through 2013 claims option; 2009 through 2013 registry option; 2011 through 2012 group practice reporting II option; and the 2012 ACO option . In the 2008 data, 53.61% of patients reported on did not meet the measure. Average performance from 2010-2014 has improved (but with wide variation) with 2014 average reported to be 95.9%.
The developer reports disparities are identified as an issue in the literature, but results for this measure's ability to detect them were not provided.
Questions for the Committee:
o Is there a gap in care that warrants a national performance measure? YES
o Is this measure "topped out" and so should be considered for Reserve Status? NO, BASED ON THE VARIATION 2010-2014
o If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare? NO
**If I am interpreting this correctly (page 3), while in 2008 there was a substantial gap, for 2012-2014 there is only a narrow (1.5-4.1%) performance gap! Based on my experience, I would expect a more substantial gap; however, if the gap is indeed this narrow, this measure should be harmonized with 0091.
**There appears to be discrepancies between the evidence reported by the developers and the CMS data reported. Depending on the data one chooses to support their review of this measure there may be a gap or the measure may have topped out. I feel a discussion of the evidence presented will be helpful.
I am unaware of evidence that disparities exist in this area of healthcare.

1c. High Priority (previously referred to as High Impact)
Comments:
**Not applicable

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. **Data source(s):** Administrative claims, Electronic Clinical Data : Registry **Specifications:** • Specifications were updated since the last review. o The numerator was edited to more closely align to the most recent evidence-based guidelines as noted in the evidence attachment. The prior numerator was: "Patients who were prescribed an inhaled bronchodilator." It has been updated to: "Patients who were prescribed a long-acting inhaled bronchodilator." o The previous denominator of "...with a diagnosis of COPD and who have an FEV1/FVC less than 60%..." had

a transcription error and has been corrected to " ...with a diagnosis of COPD (FEV1/FVC < 70%) who have an FEV1 less than 60% predicted..."

- The calculation algorithm is stated in S.18 and appears straightforward.

Questions for the Committee :

- *Are the updates to the specifications appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Initial reliability testing relied on data element-level validity testing, which per NQF guidance does not require separate reliability testing. Results from this are discussed in the following section.
-

Describe any updates to testing:

- Performance measure score-level reliability testing conducted by Mathematica Policy Research in 2012.
- Testing and analysis included 11,593,241 Medicare beneficiaries identified on claims associated with 66 PQRS GPRO groups with at least 25 eligible professionals (EPs) and 396 ACOs. Beneficiaries attributed to groups with more than 25 EPs averaged 2,974 (standard deviation = 5,105). Approximately half (52%) of the groups were attributed fewer than 1,000 beneficiaries. Beneficiaries attributed to groups with more than 100 EPs averaged 7,077 (standard deviation = 7,842).
- Groups were included if they reported at least 20 eligible cases for the measure.
- The groups were distributed across all states, the District of Columbia, Guam, and Puerto Rico.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐**No**

Method(s) of reliability testing:

- Reliability was estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error).

Results of reliability testing:

- The developer reports that although there is no universally agreed-upon minimum reliability threshold, the literature indicates reliability scores in the 0.40–0.70 range are considered moderate, and scores greater than 0.70 are considered high indicating a group's performance rates would be similar if performance were calculated on the basis of a random sample of the practice's beneficiaries.
- The reliability score for this measure was 0.85 among groups with 25 or more EPs participating in the PQRS GPRO program.
- The reliability for this measure among ACOs was not included in the analysis.

Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH)

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity was conducted by AMA-PCPI and was assessed by a 12 member expert panel
- Data element-level validity testing conducted by comparing the automated EHR report to visual inspection of the medical record.
  - The results of the data element testing conducted by comparing the automated EHR report to visual inspection of the medical record (1 site, 106 patient encounters) were:
    - Reliability: N, % Agreement, Kappa
    - Numerator: 106, 82.08%, kappa 0.1444* (0.0000-0.3248 CI)
    - Denominator: 106, 100%, kappa non-calculable** (non-calculable CI)

Describe any updates to validity testing

- The new developer, ATS, conducted new face validity testing.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

☐ **Empirical validity testing of the measure score**

Validity testing method:

- The 12-member ATS Clinical Practice Committee was asked to rate its agreement with the following statement: The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.
- The rating scale used was 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

Validity testing results:

- 88.9% of panelists agreed or strongly agreed this measure can accurately distinguish good and poor quality

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Exceptions include medical reason(s), patient reason(s) or system reason(s) for not documenting spirometry results. Although this methodology does not require the external reporting of more detailed exception data, the ATS recommends that physicians document the specific reasons for exception in patients' medical records for

purposes of optimal patient management and audit-readiness.

- Exclusion analysis was not conducted on this measure, as requested by NQF.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are the exclusions too vague? Are any patients or patient groups inappropriately excluded from the measure?*
- *Because no exclusion analyses were performed and reported by the developer, is there a threat to validity?*
- *Is Committee aware of whether the exclusions/exceptions are of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Meaningful difference analysis was not conducted on this measure as requested by NQF.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Because no analysis of meaningful differences was provided by the developer, is there a threat to validity?*

2b6. Comparability of data sources/methods:

- Not Applicable

2b7. Missing Data

- Missing data analysis was not conducted on this measure as requested by NQF

Question for the Committee:

- *Because no analysis of missing data was provided by the developer, is there a threat to validity?*

Guidance from the Validity Algorithm : 1→2→4→5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Question for the Committee:

- Are the specifications consistent with the evidence? YES

**Specifications are consistent

**The specifications are consistent with the evidence.

2a2. Reliability Testing

Comments:

**Prior data element validity: conducted by comparing the automated EHR report to visual inspection of the medical record. The results of the data element testing conducted by comparing the automated EHR report to visual inspection of the medical record (1 site, 106 patient encounters) were: Numerator: 82.08%, kappa 0.1444 (0.0000-0.3248 CI); Denominator: 100%, non-calculable kappa or CI.

New face validity by ATS Clinical Practice Committee (12 members) with 88.9% agreed or strongly agreed the measure can distinguish good and poor quality.

Questions for the Committee:

- Do the results demonstrate sufficient validity so that conclusions about quality can be made? YES
- Do you agree that the score from this measure as specified is an indicator of quality? YES

**yes

**The results demonstrate sufficient validity so that conclusions about quality can be made.

I agree the score from this measure as specified is an indicator of quality.

2b2. Validity Testing

Comments:

**2b3: Exceptions determined by assessment/interpretation of provider. Exclusion analysis was not conducted as

requested by NQF.
Questions for the Committee:
o Are the exclusions consistent with the evidence? YES
o Are the exclusions too vague? BASED ON INDIVIDUAL PRACTICE/PATIENT Are any patients or patient groups inappropriately excluded from the measure? UNCERTAIN/NO
o Because no exclusion analyses were performed and reported by the developer, is there a threat to validity? NO
o Is Committee aware of whether the exclusions/exceptions are of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? NO
2b4: No risk adjustment--NA
2b5: Meaningful difference analysis not conducted on NQF recommendation.
Question for the Committee:
o Does this measure identify meaningful differences about quality? NA
o Because no analysis of meaningful differences was provided by the developer, is there a threat to validity? No
2b6: Not applicable
2b7: No missing data analysis conducted on NQF recommendation.
Question for the Committee:
o Because no analysis of missing data was provided by the developer, is there a threat to validity? NO
Guidance from the Validity Algorithm : 1→2→4→5 (highest eligible rating is MODERATE)
**Missing data analysis not performed
**Exclusions: The exclusions are appropriate.
Meaningful Differences: Depending on the evidence used to determine the current performance on this measure, there appears to be potential to identify meaningful differences in patient care.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Performance measure score-level reliability testing conducted by Mathematica Policy Research in 2012.
Testing/analysis included 11 million + Medicare beneficiariesID'd on claims associated with 66 PQRS GPRO groups with at least 25 eligible professionals (EPs) and 396 ACOs. Approximately half (52%) of the groups were attributed fewer than 1,000 beneficiaries. Groups were included if they reported at least 20 eligible cases for the measure. Geographic representation was widespread including all 50 states, DC, Guam and Puerto Rico.
The reliability score for this measure was 0.85 (considered high) among groups with 25 or more providers, but the reliability for this measure among ACOs was not included in the analysis.
Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH)
Questions for the Committee:
o Do the results demonstrate sufficient reliability so that differences in performance can be identified? YES
**yes
**The results demonstrate sufficient reliability so that differences in performance can be identified.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:
o *Are the required data elements routinely generated and used during care delivery?*
o *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care.
Questions for the Committee:
o Are the required data elements routinely generated and used during care delivery? YES
o Are the required data elements available in electronic form, e.g., EHR or other electronic sources? YES
**Primary care physicians managing COPD patients may not perform spirometry. Would such cases be included in this measure?
**This measure appears to be feasible because •all data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
- This measure is planned for integration into the CMS Physician Compare Program. Although Physician Compare has been launched, this measure is not yet included as of 12/14/15.
- This measure has been in use for the CMS PQRS program since 2007.

Improvement results : The developer referenced the 2011-2014 PQRS performance data to present improvement results.

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms : The developer did not identify any unintended consequences related to this measure.

Feedback : No feedback provided on QPS. Measure reviewed by MAP for Physician Quality Reporting System (PQRS) in 2012, 2013, and 2014. The measure also was reviewed in Physician Compare and Value-Based Payment Modifier Program in 2014. MAP recommended the developer explore creating a composite of all COPD measures and then link that composite with the COPD resource use measure.

***Questions for the Committee**:*
o *Are you aware of any implementation issues with this measure?*
o *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments
Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement
4c. Unintended Consequences
Comments:
**Currently the measure is publicly reported and used in an accountability program. It is planned for integration into the CMS Physician Compare program but has not yet been included as of 12/14/15.
No unexpected findings were reported during implementation.
Questions for the Committee:
o Are you aware of any implementation issues with this measure? NO
o Do the benefits of the measure outweigh any potential unintended consequences? YES
**Benefits outweigh any unintended consequences for a process measure for performance of spirometry for COPD. MODERATE
**I am unaware of any implementation issues with this measure and therefore it appears to have high usability. The benefits of the measure (ensuring patients receive the recommended medication) outweigh any potential unintended consequences.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2856: Pharmacotherapy Management of COPD Exacerbation

Harmonization:

- No harmonization plan mentioned by developer

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM

NOTE: As instructed, responses from the 2012 comprehensive review evidence form submitted by the AMS-PCPI are changed to red font. Items in black font within a response represent minor edits reflecting 2015 updates to GOLD.

NQF #: 0102 NQF Project: Pulmonary Project

1. IMPACT, OPPORTUITY, 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)

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)*:

The measure focus is the process of prescribing an inhaled bronchodilator to patients with a diagnosis of COPD and

symptoms (eg, dyspnea, cough/sputum, wheezing). This process is directly related to managing COPD symptoms, as well as reducing long-term lung function decline, COPD exacerbations, inpatient hospitalizations, and mortality. Inhaled bronchodilator therapy is effective in treating and managing the symptoms of COPD and improving quality of life, particularly for those patients with moderate to very severe COPD, which should lead to less comorbid disease, physical dysfunction, and death from COPD.

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

Clinical Practice Guideline

Systematic review of body of evidence (other than within guideline development)

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)*:

The evidence cited for this measure is directly related to the usefulness of prescribing an inhaled bronchodilator for adults with COPD and symptoms. There are no differences from the measure focus and measure target population.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles*): The quantity of studies reviewed in the ACP/ACCP/ATS/ERS guideline was not explicitly stated, but the guideline paper references 62 articles, and there is some specific information regarding evidence for bronchodilator use. This guideline is based on a targeted literature update from March 2007 to December 2009 to evaluate the evidence and update the 2007 ACP clinical practice guideline on diagnosis and management of stable COPD. Pooled results from 9 long-term trials, some of which were not statistically significant, demonstrated that inhaled therapies (long-acting bronchodilators, inhaled corticosteroids, or combination bronchodilator and corticosteroid therapy) reduced the annual decline in mean FEV1 more than placebo did. Monotherapy trials reported absolute decreases in the annual rate of FEV1 decline associated with use of tiotropium (40 mL/y), inhaled corticosteroids (44 mL/y), and long-acting ß-agonists (42 mL/y). Other studies have demonstrated that combinations of inhaled agents are not more effective than monotherapy for slowing declines in lung function. (Qaseem et al, 2011)

The quantity of studies reviewed in the GOLD report was not stated, but the paper references 503 articles, and there is some specific information regarding Evidence Category A:

Sources of Evidence: Randomized controlled trials (RCTs). Rich body of data.

Definition: Evidence is from endpoints of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants. (GOLD 2015)

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):* The ACP/ACCP/ATS/ERS guideline recommendation was graded as a strong recommendation, with moderate-quality evidence. A strong recommendation means that benefits clearly outweigh risks and burden, or risks and burden clearly outweigh benefits. Evidence is considered moderate quality when it is obtained from RCTs with important limitations—for example, biased assessment of the treatment effect, large loss to follow-up, lack of blinding, unexplained heterogeneity (even if it is generated from rigorous RCTs), indirect evidence originating from similar (but not identical) populations of interest, and

RCTs with a very small number of participants or observed events. In addition, evidence from well-designed controlled trials without randomization, well designed cohort or case–control analytic studies, and multiple time series with or without intervention are in this category. Moderate-quality evidence also means that further research will probably have an important effect on our confidence in the estimate of effect and may change the estimate. (Amir Qaseem, MD, PhD, MHA; Vincenza Snow, MD; Douglas K. Owens, MD, MS; and Paul Shekelle, MD, PhD, for the Clinical Guidelines Committee of the American College of Physicians. The Development of Clinical Practice Guidelines and Guidance Statements of the American College of Physicians: Summary of Methods. Ann Intern Med. 2010;153:194-199.)

The quality of studies reviewed in the GOLD report was not addressed, but there is some specific information regarding Evidence Category A:

Sources of Evidence: Randomized controlled trials (RCTs). Rich body of data.

Definition: Evidence is from endpoints of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants. (GOLD 2015)

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The ACP/ACCP/ATS/ERS guideline recommendation for inhaled bronchodilators is not consistently recommended for all COPD populations. Rather, the guideline explains that updated evidence reconfirms prior findings that the patients who benefit the most from inhaled therapies (anticholinergics, long-acting ß-agonists, or corticosteroids) are those who have respiratory symptoms and airflow obstruction with FEV1 less than 60% predicted. Although some patients who were studied had an FEV1 greater than 60% predicted, the mean FEV1 of the included patients has been 60% predicted or less for most COPD treatment trials. Among symptomatic patients with FEV1 greater than 50% predicted but less than 80% predicted or those with normal airflow but who have chronic sputum production (at-risk individuals), 7 large studies of inhaled corticosteroids or short- or long-acting anticholinergics that lasted at least 1 year (including 2 published since the 2007 review) found little to no improvement in exacerbations, health-related quality of life, COPD hospitalizations, or mortality. Therefore, there is limited and conflicting evidence of health benefits resulting from initiation of inhaled bronchodilators (anticholinergics or long-acting ß-agonists) in symptomatic patients with FEV1 between 60% and 80% predicted as documented by spirometry. Individual patients may benefit from the therapy and may show improvement in their respiratory symptoms. However, the duration of maintenance therapy and the frequency of reevaluation once a patient is receiving therapy are unknown because evidence is limited. Further research is needed to evaluate the health benefits of inhaled therapies (anticholinergics or long-acting ß-agonists) in symptomatic patients with FEV1 between 60% and 80% predicted. The mean FEV1 was less than 60% predicted in the majority of the trials that evaluated the management of COPD. Monotherapy with a long-acting inhaled ß-agonist or a long-acting inhaled anticholinergic is beneficial in reducing exacerbations and improving health-related quality of life. Evidence was inconclusive regarding the effect of inhaled agents (anticholinergics and long-acting ß-agonists) on mortality, hospitalizations, and dyspnea. This recommendation does not address the occasional use of short-acting inhaled bronchodilators for acute symptom relief. (Qaseem et al, 2011)

The consistency of studies reviewed in the GOLD report was not addressed, but there is some specific information regarding Evidence Category A:

Sources of Evidence: Randomized controlled trials (RCTs). Rich body of data.

Definition: Evidence is from endpoints of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants. (GOLD 2015)

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

benefit - benefit over harms):

The ACP/ACCP/ATS/ERS guideline panel included representatives from each of the 4 collaborating organizations, and the resulting guideline represents an official and joint clinical practice guideline from those organizations. The guideline panel communicated via conference calls and e-mails. The members reached agreement and resolved any disagreements through facilitated discussion. The final recommendations were approved by unanimous vote.

The net benefit of studies reviewed in the GOLD report was not addressed, but there is some specific information regarding Evidence Category A:

Sources of Evidence: Randomized controlled trials (RCTs). Rich body of data.

Definition: Evidence is from endpoints of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants. (GOLD 2015)

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The ACP/ACCP/ATS/ERS guideline panel included representatives from each of the 4 collaborating organizations, and the resulting guideline represents an official and joint clinical practice guideline from those organizations. Potential Conflicts of Interest: Any financial and nonfinancial conflicts of interest of the group members were declared, discussed, and resolved. Dr. Wilt: Grant: American College of Physicians; Payment for manuscript preparation: American College of Physicians. Dr. Hanania: Consultancy: GlaxoSmithKline, Boehringer Ingelheim, Novartis, Pfizer, Sunovion, Pearl, Forest; Grants/grants pending (money to institution): GlaxoSmith-Kline, Boehringer Ingelheim, Novartis, Pfizer, Sunovion; Payment for lectures including service on speakers bureaus: GlaxoSmithKline, Astra-Zeneca, Boehringer Ingelheim, Merck. Dr. Criner: Consultancy: Uptake Medical, PortAero, Pulmonx; Grants/grants pending (money to institution): Aeris Therapeutics, Emphysas Medica. Dr. van der Molen: Consultancy: MSD, AstraZeneca, GlaxoSmithKline, Nycomed; Grants/grants pending (money to institution): AstraZeneca, GlaxoSmithKline, Novartis; Payment for lectures inclucing service on speakers bureaus: AstraZeneca, Nycomed, GlaxoSmithKline, MSD. Dr. Marciniuk: Board membership: American College of Chest Physicians, Chest Foundation, Lung Association of Saskatchewan, Canadian COPD Alliance, Canadian Thoracic Society; Consultancy (no payment received): Public Health Agency of Canada, Canadian Agency for Drugs and Technology in Health; Consultancy: Saskatchewan Medical Association; Consultancy (money to institution): AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Saskatchewan

Health Quality Council, Novartis, Nycomed, Pfizer; Employment: University of Saskatchewan, Saskatoon Health Region; Grants/grants pending (money to institution): Canadian Institute of Health Research, AstraZeneca, GlaxoSmithKline, Lung Association of Saskatchewan, Nycomed, Pfizer, Novartis, Saskatchewan Health Research Foundation, Schering-Plough, Saskatchewan Ministry of Health; Payment for lectures including service on speakers bureaus: AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Pfizer, Lung Association of Saskatchewan, Canadian Thoracic Society, American Thoracic Society. Dr. Wedzicha: Grants/grants pending (money to institution): Boehringer Ingelheim; Board membership: GlaxoSmithKline, Novartis, Bayer, Pfizer, Medimmune/Astra-Zeneca, Danone/Nutricia, Nycomed; Consultancy: Chiesi; Consultancy (money to institution): Novartis; Grants/grants pending (money to institution):

GlaxoSmithKline, Novartis, Chiesi, AstraZeneca, Johnson & Johnson; Payment for lectures including service on speakers bureaus: Boehringer Ingelheim, GlaxoSmithKline, Pfizer, Bayer, Nycomed, Chiesi; Travel/accommodations/meeting expenses unrelated to activities listed: Boehringer Ingelheim. Dr. Shekelle: Employment: Veterans Affairs Medical Center; Grants/grants pending (money to institution): Agency for Healthcare Research and Quality, National Institutes of Health, Veterans Administration; Royalties: UpToDate; Travel/accommodations/meetings expenses unrelated to activities listed: Travel to meetings sponsored by AHRQ, the Health Foundation, the University of Michigan, VA, Italian regional health authority, and RAND. Disclosures can also be viewed at

www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M11-0925.

The GOLD Science Committee includes physician representatives from around the globe, including Denmark, Spain, Italy, England, Germany, Japan, Canada, Washington, Michigan, and Texas; disclosure forms are posted on the GOLD website, www.goldcopd.org.

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

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: Moderate

1c.14 Summary of Controversy/Contradictory Evidence: Adverse effects related to inhaled long-acting anticholinergics or long-acting ß-agonists range from mild (for example, dry mouth) to potentially serious (for example, cardiovascular events). Pooled analyses of results from trials of monotherapy show no statistically significant differences in outcomes among various monotherapies. However, some of the large recent trials have shown that different monotherapies may have a greater effect on certain outcomes. These observed effects need to be confirmed with further comparative effectiveness studies. Clinicians should base selection of treatment from among various monotherapies on individual patient preferences, cost, and adverse effect profile. (Qaseem et al, 2011)

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

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

Recommendation 3: For stable COPD patients with respiratory symptoms and FEV1 <60% predicted, ACP, ACCP, ATS, and ERS recommend treatment with inhaled bronchodilators (Grade: strong recommendation, moderate-quality evidence).

Recommendation 4: ACP, ACCP, ATS, and ERS recommend that clinicians prescribe monotherapy using either long-acting inhaled anticholinergics or long-acting inhaled ß-agonists for symptomatic patients with COPD and FEV1 <60% predicted. (Grade: strong recommendation, moderate-quality evidence). Clinicians should base the choice of specific monotherapy on patient preference, cost, and adverse effect profile. Monotherapy with a long-acting inhaled agent (long-acting anticholinergic, long-acting ß-agonist, or corticosteroid) was superior to placebo or short-acting anticholinergic therapy in reducing exacerbations (Qaseem et al, 2011)

Bronchodilator medications are given on either an as-needed basis or a regular basis to reduce or prevent symptoms (Evidence A). Bronchodilator medications are central to symptom management in COPD. Inhaled therapy is preferred. Long-acting inhaled bronchodilators are convenient and more effective at producing maintained symptom relief than short-acting bronchodilators. Based on efficacy and side effects, inhaled bronchodilators are preferred over oral bronchodilators. (Evidence A). For both $beta_2$-agonists and anticholinergics, long-acting formulations are preferred over short-acting formulations (Evidence A) (GOLD 2015).

1c.17 Clinical Practice Guideline Citation: Amir Qaseem, MD, PhD, MHA; Timothy J. Wilt, MD, MPH; Steven E. Weinberger, MD; Nicola A. Hanania, MD, MS; Gerard Criner, MD; Thys van der Molen, PhD; Darcy D. Marciniuk, MD; Tom

Denberg, MD, PhD; Holger Schunemann, MD, PhD, MSc; Wisia Wedzicha, PhD; Roderick MacDonald, MS; and Paul Shekelle, MD, PhD, for the American College of Physicians, the American College of Chest Physicians, the American Thoracic Society, and the European Respiratory Society. Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Ann Intern Med. 2011;155:179-191.

Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015. Available at http://www.goldcopd.org.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/content.aspx?id=34205&search=copd

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: See 1.c.10

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

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: Strong

1c.24 Rationale for Using this Guideline Over Others: It is the ATS policy to use guidelines, which are evidence-based, applicable to physicians and other health-care providers, and developed by a national specialty organization or government agency.

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: Moderate 1c.26 Quality: Moderate1c.27 Consistency: Moderate

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0102_Evidence_2015.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Despite major efforts to broadly disseminate the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines and use of COPD performance measures across different specialty societies, management of COPD, and specifically prescription for long-acting inhaled bronchodilators, remains suboptimal. Studies show a wide range of deficiencies in adherence to guidelines regarding long-acting inhaled bronchodilator use across different settings (Asche et al., 2012; CDC, 2012; Fitch, et al., 2011; Nantsupawat et al., 2012; Perez et al., 2011; Sharif, et al., 2013). Underuse of broncholidators were found related to hospital readmissions and to increased total costs of services when compared to patient care adhering to GOLD guidelines (Asche et al., 2012; Nantsupawat et al., 2012).

Suboptimal COPD management has implications for severity of illness, disease progression, patient quality of life and health status, exacerbations (and associated costs) and mortality. Improved adherence to COPD management guidelines, specifically appropriate use of long-acting inhaled bronchodilators, has the potential to improve clinical outcomes and cost of care related to COPD. As a result, we believe this measure will continue to increase appropriate long-acting inhaled bronchodilator use, improving patient management and total costs of COPD.

Citations:

Asche CV, Leader S, Plauschinat C, Raparla S, Yan M, Ye X, Young D. Adherence to current guidelines for chronic obstructive pulmonary disease (COPD) among patients treated with combination of long-acting bronchodilators or inhaled corticosteroids. Int J Chron Obstruct Pulmon Dis. 2012;7:201-9.

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Fitch K, Iwasaki K, Pvenson B. Plauschinat C, Zhang J. Variation in adherence with Global Initiative for Chronic Obstructive Lung Disease (GOLD) drug therapy guidelines: a retrospective actuarial claims data analysis. Curr Med Res Opin. 2011 Jul;27(7):1425-9.

Nantsupawat T, Limsuwat C, Nugent K. Factors affecting chronic obstructive pulmonary disease early rehospitalization. Chron Respir Dis. 2012 May;9(2):93-8.

Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.

Sharif R, Cuevas CR, Wang Y, Arora M, Sharma G. Guideline adherence in management of stable chronic obstructive pulmonary disease. Respir Med. 2013 Jul;107(7):1046-52.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
According to a study analyzing the quality of health care in the United States, on average, patients with COPD received the recommended care at an aggregate rate (based on 20 quality indicators) of 58 percent (McGlynn et al., 2003).

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines recommend inhaled bronchodilators as a cornerstone of COPD symptom management;
however, PCPs often turn to other agents as first-line COPD therapy (Barr et al, 2005; Foster et al, 2007).

A cross-sectional study implemented in July 2008 was designed to assess attitudes and barriers to COPD guideline usage. Five hundred U.S. PCPs (309 family medicine physicians, 191 internists) were included in the analysis. 78.4% of PCPs agreed that a long-acting bronchodilator (LABD) should be added for patients with stage 2–3 COPD whose dyspnea during daily activities is not relieved with an as-needed short-acting bronchodilator. However, only 25.8% of the PCPs “nearly always” recommend using an LABD daily for

patients with COPD and mild exertional dyspnea (Salinas et al, 2011).

In a recent study of general medicine practices, 154 clinicians completed a survey to identify barriers to implementing seven recommendations from the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines. Adherence was only 54% to prescribing long-acting bronchodilators when FEV(1) <80% predicted (Perez, et al, 2011).

Another study of 615 patients being treated for COPD by a general practitioner, less than half the patients in all groups used short-acting bronchodilators. Prescribing long-acting bronchodilators or inhaled corticosteroids conformed to GOLD guidelines in two-thirds of patients with GOLD stage III or IV disease, and approximately half of the less severe patients (Jochmann et al, 2010).

CMS Physician Quality Reporting Initiative/System:
This measure was used in the CMS Physician Quality Reporting Initiative/System (PQRI/S) in the: 2007 through 2011 claims option; 2009 through 2011 registry option; and the 2011 group practice reporting II option.

There is a gap in care as shown by this 2008 data; 53.61% of patients reported on did not meet the measure.(1)

10th percentile: 0.00%
25th percentile: 1.75%
50th percentile: 30.77%
75th percentile: 74.07%
90th percentile: 89.44%

Exception rate: 9.98%This measure has been in use by the CMS Physician Quality Reporting Initiative/System (PQRI/S) since 2007 with the following reporting options:
•2007 – Claims option
•2008-2010, 2013 – Claims and registry options
•2011 – Claims, registry and GPRO II options
•2012 – Claims, registry, GPRO II and ACO options

Average performance rate:
2010 - 89.7%
2011 - 73.4%
2012 - 98.5%
2013 - 97.0%
2014 - 95.9%

Data from CMS (1) indicates a favorable overall trend 2010-2014. Most recent data indicate a 4% gap in care for 2014. This gap is not aligned with research findings cited in 1b.3.

Based on our evaluation of performance data, recent evidence-based guidelines recommending the use of long-acting bronchodilators among patients with COPD and FEV1 <60% predicted, and research that demonstrates a lack of adherence to these guidelines (see 1b.3), the ATS Subcommittee on COPD Measures recommends an update to the specifications of this measure as described in this comprehensive review submission. The updated measure will be released in the PQRS 2017 performance year. We anticipate a reduction in performance rates until the guidelines on long-acting inhaled bronchodilators become broadly adopted. We believe the updated measure will result in further improvement in COPD patient care quality.

(1) Source: Timothy Jackson, CMS.

_____

Performance scores from 2012 comprehensive review submitted by PCPI to provide history.

CMS Physician Quality Reporting Initiative/System:

This measure was used in the CMS Physician Quality Reporting Initiative/System (PQRI/S) in the: 2007 through 2011 claims option; 2009 through 2011 registry option; and the 2011 group practice reporting II option.

There is a gap in care as shown by this 2008 data; 53.61% of patients reported on did not meet the measure.(1)

10th percentile: 0.00%
25th percentile: 1.75%
50th percentile: 30.77%
75th percentile: 74.07%
90th percentile: 89.44%

Exception rate: 9.98%

(1) Confidential CMS PQRI Performance Information by Measure. Jan-Sept TAP file.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Studies consistently show suboptimal adherence to guidelines specifically to prescribing inhaled bronchodilator. A retrospective study of 364 patients at Geisinger Health Plan showed adherence to guidelines for bronchodilators ranged from 35%-46% (Asche et al., 2012). A multiyear survey completed by the North Carolina state health department (in collaboration with CDC) found only 48.1% of the 2,187 persons with COPD reported daily use of COPD medication (CDC, 2012). A study of 450 COPD patients conducted at an academic medical center showed 54.7% received treatment according to guidelines (Sharif, et al., 2013). A study of 81 patients hospitalized for COPD exacerbations found 32% were discharged without prescribed long-acting bronchodilators and/or inhaled corticosteroids (Nantsupawat et al., 2012).

A large, retrospective actuarial claims data analysis of 44,366 cases showed "claims for short acting bronchodilator therapy without concomitant use of long acting bronchodilators were identified for 20% of moderate, 14% of severe and 8% of very severe COPD patients; and claims for single long acting bronchodilator therapy in combination with inhaled corticosteroid therapy were identified for 12% of moderate, 19% of severe and 2% of very severe COPD patients" (Fitch, et al., 2011). A survey of 154 clinicians found prescription of inhaled bronchodilators according to guidelines was 54% when FEV1 <80% predicted (Perez et al., 2011).

Citations:

Asche CV, Leader S, Plauschinat C, Raparla S, Yan M, Ye X, Young D. Adherence to current guidelines for chronic obstructive pulmonary disease (COPD) among patients treated with combination of long-acting bronchodilators or inhaled corticosteroids. Int J Chron Obstruct Pulmon Dis. 2012;7:201-9.

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Fitch K, Iwasaki K, Pvenson B. Plauschinat C, Zhang J. Variation in adherence with Global Initiative for Chronic Obstructive Lung Disease (GOLD) drug therapy guidelines: a retrospective actuarial claims data analysis. Curr Med Res Opin. 2011 Jul;27(7):1425-9.

Nantsupawat T, Limsuwat C, Nugent K. Factors affecting chronic obstructive pulmonary disease early rehospitalization. Chron Respir Dis. 2012 May;9(2):93-8.

Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.

Sharif R, Cuevas CR, Wang Y, Arora M, Sharma G. Guideline adherence in management of stable chronic obstructive pulmonary disease. Respir Med. 2013 Jul;107(7):1046-52.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities*

include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.
We are not aware of any disparities data from this measure as specified. Please see 1b.5 for a summary of our findings in the literature regarding disparities.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Studies have been done to show associations between education level and income and outcomes related to COPD (Eisner et al., 2011; Holt et al., 2011). Studies also show association between gender and race on the incidence/severity of COPD (Bruse et al., 2011; Diaz et al., 2014; Foreman et al., 2011; Han et al., 2011). However, we found limited research regarding potential disparities on inhaled bronchodilator use in the US.

One cross-sectional study conducted in the UK examined differences in COPD management in three multiethnic, socially deprived communities. Black patients with COPD were less likely to have dyspnea, less likely to be prescribed inhaled bronchodilators and less likely to be referred to pulmonary rehabilitation programs. This group was also more likely to be hospitalized for respiratory conditions on one of the three communities. South Asians were also less likely to have dyspnea and be referred for pulmonary rehabilitation. However, they received medication similar to the white population and had similar hospitalization rates. These differences in medication use are not likely attributable to access or insurance as all patients were covered by the National Health Service (Martin et al., 2012).

One study on adherence to COPD guidelines conducted at an urban academic center in Texas found no association between age, sex, and race and guideline adherence (Sharif, et al., 2013).

The ATS is aware of health disparities related to respiratory diseases and has recently created a Health Equality Subcommittee of the Health Policy Committee. This group has been tasked with providing recommendations for moving toward respiratory health equality to include improving environmental factors, healthy lifestyle promotion, high quality healthcare (prevention, screening, diagnosis and treatment) and further research (Celedón et al., 2014).

Citations:

Bruse S, Sood A, Petersen H, Liu Y, Leng S, Celedón JC, Gilliland F, Celli B, Belinsky SA, Tesfaigzi Y. New Mexican Hispanic smokers have lower odds of chronic obstructive pulmonary disease and less decline in lung function than non-Hispanic whites. Am J Respir Crit Care Med. 2011 Dec 1;184(11):1254-60.

Celedón JC, Roman J, Schraufnagel DE, Thomas A, Samet J. Respiratory health equality in the United States. The American thoracic society perspective. Ann Am Thorac Soc. 2014 May;11(4):473-9.

Diaz AA, Come CE, Mannino DM, Pinto-Plata V, Divo MJ, Bigelow C, Celli B, Washko GR. Obstructive lung disease in Mexican Americans and non-Hispanic whites: an analysis of diagnosis and survival in the National Health and Nutritional Examination Survey III Follow-up Study. Chest. 2014 Feb;145(2):282-9.

Eisner MD, Blanc PD, Omachi TA, Yelin EH, Sidney S, Katz PP, Ackerson LM, Sanchez G, Tolstykh I, Iribarren C. Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health 2011;65:26–34.

Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, et al. Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene Study. Am J Respir Crit Care Med. 2011 Aug 15;184(4):414-20.

Han MK, Curran-Everett D, Dransfield MT, Criner GJ, Zhang L, Murphy JR, Hansel NN, DeMeo DL, Hanania NA, Regan EA, Make BJ, Martinez FJ, Westney GE, Foreman MG; COPDGene Investigators. Racial differences in quality of life in patients with COPD. Chest. 2011 Nov;140(5):1169-76.

Holt JB, Zhang X, Presley-Cantrell L, Croft JB. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. Int J Chron Obstruct Pulmon Dis. 2011; 6 321–328.

Martin A, Badrick E, Mathur R, Hull S. Effect of ethnicity on prevalence, severity, and management of COPD in general practice. Br J Gen Pract. 2012 Feb;62(595):e76-81.

Sharif R, Cuevas CR, Wang Y, Arora M, Sharma G. Guideline adherence in management of stable chronic obstructive pulmonary disease. Respir Med. 2013 Jul;107(7):1046-52.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

A 2013 analysis estimates the COPD prevalence in US adults aged 40-79 to be 14-15% (Tiler et al., 2013). Significant mortality and morbidity and healthcare costs are associated with COPD.

COPD is the third ranked cause of death in the US and the number of deaths due to COPD increased 3.5% between 2010 and 2011 (CDC, 2012; Heron, 2015).

Worldwide, the Global Burden of Disease estimates 328 million people with COPD including 168 million men and 160 million women. COPD was the sixth ranked cause of death in 1990, moving upward to fourth in 2000 and estimated to be the third leading cause of death globally by 2020 (Lopez-Campos et al., 2015). The World Health Organization estimates 3 million people with COPD in the world die each year due to COPD (Diaz-Guzman and Mannino, 2014). COPD exacerbations resulting in hospitalizations correlate with 1-year mortality of 21% and 5-year mortality of 55% (Lopez-Campos et al., 2015).

Disability as a result of COPD is high globally with 29.4 million [Years Lost to Disability] YLD. YLD due to COPD rankings have moved from sixth to fifth from 1990 to 2010 (Lopez-Campos et al., 2015).

COPD exacerbations impact quality of life and healthcare costs. The cost of hospitalizations related to COPD exacerbations in the US is estimated to be $18 billion annually (Lopez-Campos et al., 2015). Thirty day readmission rates for patients discharged from hospitals for COPD exacerbations are 20% and 35% at 90-days (Prieto Centurion et al., 2012). Several studies show that COPD management according to guidelines prevents exacerbation and hospital admission (Asche, et al., 2012; Martin et al., 2012).

Mathematica Policy Research conducted an analysis on CMS data in 2012 finding that per capita costs for beneficiaries with COPD was $24,901, compared to overall per capita costs of $10,734 (Mathematica Policy Research, 2014).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Citations:

Asche CV, Leader S, Plauschinat C, Raparla S, Yan M, Ye X, Young D. Adherence to current guidelines for chronic obstructive pulmonary disease (COPD) among patients treated with combination of long-acting bronchodilators or inhaled corticosteroids. Int J Chron Obstruct Pulmon Dis. 2012;7:201-9.

Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.

Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clin Chest Med. 2014 Mar 35(1):7-16.

Heron M. Deaths: leading causes for 2011. Natl Vit Stat Rep. 2015 Jul 27;64(7):1-96

Lopez-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology. 2015 Oct 23. [Epub ahead of print].

Martin A, Badrick E, Mathur R, Hull S. Effect of ethnicity on prevalence, severity, and management of COPD in general practice. Br J Gen Pract. 2012 Feb;62(595):e76-81.

Mathematica Policy Research. Experience Report for the Performance Year 2012 Quality and Resource Use Reports. January 8, 2014. Accessed December 7, 2015. Accessible at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2012-QRUR_Experience_Report.pdf

Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.

Tiler T, Dillon C, Paulose-Ram R, Hnizdo E, Doney B. Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post- bronchodilator spirometry: the National Health and Nutrition Examination Survey (NHANES) 2007-2010. Respir Res. 2013 Oct 9;14:103.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
The specifications are included in this form.

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
This measure includes edits to correct a prior transcription error.

The incorrect denominator was: ...with a diagnosis of COPD and who have an FEV1/FVC less than 60%...

The corrected denominator is: ...with a diagnosis of COPD (FEV1/FVC < 70%) who have an FEV1 less than 60% predicted...

This measure also includes edits to more closely align to the most recent evidence-based guidelines as noted in the evidence attachment.

The prior numerator included: Patients who were prescribed an inhaled bronchodilator.

The updated numerator includes: Patients who were prescribed a long-acting inhaled bronchodilator.

The updated measure is planned for release in the 2017 PQRS performance year.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed a long-acting inhaled bronchodilator

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once per reporting period

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Definition:
Prescribed – Includes patients who are currently receiving medication(s) that follow the treatment plan recommended at an encounter during the reporting period, even if the prescription for that medication was ordered prior to the encounter.

NUMERATOR NOTE: The correct combination of numerator code(s) must be reported on the claim form in order to properly report this measure. The "correct combination" of codes may require the submission of multiple numerator codes.

Numerator Quality-Data Coding Options for Reporting Satisfactorily:

Patient Prescribed Long-acting Inhaled Bronchodilator Therapy
(One CPT II code & one quality-data code [4025F & G8924] are required on the claim form to submit this numerator option)

Performance Met:
CPT II 4025F: Long-acting inhaled bronchodilator prescribed (NOTE: pending edited CPT II code)

AND

G8924: Spirometry test results demonstrate FEV1/FVC < 70%, FEV1 < 60% predicted and patient has COPD symptoms (eg, dyspnea, cough/sputum, wheezing) (NOTE: CMS approved edited G-code for 2017 PQRS year)

OR

Patient not Documented to have Long-acting Inhaled Bronchodilator Prescribed for Medical, Patient, or System Reasons
(One CPT II code & one quality-data code [4025F-xP & G8924] are required on the claim form to submit this numerator option)
Append a modifier (1P, 2P or 3P) to CPT Category II code 4025F to report documented circumstances that appropriately exclude patients from the denominator.
Medical Performance Exclusion, Patient Performance Exclusion, or System Performance

Exclusion:
4025F with 1P: Documentation of medical reason(s) for not prescribing an inhaled bronchodilator (e.g., contraindication due to comorbidities)
4025F with 2P: Documentation of patient reason(s) for not prescribing an inhaled bronchodilator
4025F with 3P: Documentation of system reason(s) for not prescribing an inhaled bronchodilator (e.g., not covered by insurance)

AND

G8924: Spirometry test results demonstrate FEV1/FVC < 70%, FEV1 < 60% predicted and patient has COPD symptoms (eg, dyspnea, cough/sputum, wheezing)

OR

If patient is not eligible for this measure because spirometry results demonstrate FEV1/FVC >= 70% or FEV1 >= 60% predicted or patient does not have COPD symptoms, report:
Spirometry Results Demonstrate FEV1/FVC >= 70% or FEV1 >= 60% or Patient does not have COPD symptoms
(One quality-data code [G8925 or G8926] is required on the claim form to submit this numerator option)

Other Performance Exclusion: G8925: Spirometry test results demonstrate FEV1/FVC >= 70% or FEV1 >= 60% predicted or patient does not have COPD symptoms

OR

Spirometry Test not Performed or Documented
Other Performance Exclusion: G8926: Spirometry test not performed or documented, reason not given

OR

Patient not Documented to have Long-acting Inhaled Bronchodilator Prescribed, Reason not Otherwise Specified
(One CPT II code & one quality-data code [4025F-8P & G8924] are required on the claim form to submit this numerator option)
Append a reporting modifier (8P) to CPT Category II code 4025F to report circumstances when the action described in the numerator is not performed and the reason is not otherwise specified.

Performance Not Met:

4025F with 8P: Long-acting inhaled bronchodilator not prescribed, reason not otherwise specified

AND

G8924: Spirometry test results demonstrate FEV1/FVC < 70%, FEV1 < 60% predicted and patient has COPD symptoms (eg, dyspnea, cough/sputum, wheezing)

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
All patients aged 18 years and older with a diagnosis of COPD, who have FEV1/FVC < 70%, FEV1 <60% predicted and have symptoms (eg, dyspnea, cough/sputum, wheezing)

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
For Claims:
Patients aged >= 18 years on date of encounter

AND

J41.0, J41.1, J41.8, J42, J43.0, J43.1, J43.2, J43.8, J43.9, J44.0, J44.1, J44.9

AND
Patient encounter during the reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
ATS continues to use the PCPI exception methodology that uses three categories of exception reasons for which a patient may be removed from the denominator of an individual measure: medical, patient and system reasons.

Exceptions are used to remove patients from the denominator of a performance measure when a patient does not receive a therapy or service AND that therapy or service would not be appropriate due to specific reasons; otherwise, the patient would meet the denominator criteria. Exceptions are not absolute, and the application of exceptions is based on clinical judgment, individual patient characteristics, or patient preferences. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions include medical reason(s), patient reason(s) or system reason(s) for not prescribing long-acting inhaled bronchodilators. Although this methodology does not require the external reporting of more detailed exception data, the ATS recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
For Claims:

Patient not Documented to have Long-acting Inhaled Bronchodilator Prescribed for Medical, Patient, or System Reasons
(One CPT II code & one quality-data code [4025F-xP & G8924] are required on the claim form to submit this numerator option)
Append a modifier (1P, 2P or 3P) to CPT Category II code 4025F to report documented circumstances that appropriately exclude patients from the denominator.
Medical Performance Exclusion, Patient Performance Exclusion, or System Performance Exclusion:
4025F with 1P: Documentation of medical reason(s) for not prescribing a long-acting inhaled bronchodilator, e.g., contraindicated due to comorbidities
OR
4025F with 2P: Documentation of patient reason(s) for not prescribing a long-acting inhaled bronchodilator
OR
4025F with 3P: Documentation of system reason(s) for not prescribing a long-acting inhaled bronchodilator, e.g., not covered by insurance

AND

G8924: Spirometry test results demonstrate FEV1/FVC < 70%, FEV1 < 60% predicted and patient has COPD symptoms (e.g., dyspnea, cough/sputum, wheezing)

NOTE: CMS approved edited G-code for 2017 PQRS year and edited CPT II code is pending

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
We encourage the results of this measure to be stratified by race, ethnicity, primary language, and administrative sex.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific*

Acceptability)
No risk adjustment or risk stratification.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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.)*
NOTE: This sequence of steps has not been edited to reflect updated CPT II or G-codes. It will be edited once all updated CPT II or G-codes are finalized.

1. Start with Denominator

2. Check Patient Age:
a. If the Age is greater than or equal to 18 years of age on Date of Service and equals No during the measurement period, do not include in Eligible Patient Population. Stop Processing.
b. If the Age is greater than or equal to 18 years of age on Date of Service and equals Yes during the measurement period, proceed to check Patient Diagnosis.

3. Check Patient Diagnosis:
a. If Diagnosis of COPD as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
b. If Diagnosis of COPD as Listed in the Denominator equals Yes, proceed to check Encounter Performed.

4. Check Encounter Performed:
a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
b. If Encounter as Listed in the Denominator equals Yes, include in the Eligible population.

5. Denominator Population:
a. Denominator population is all Eligible Patients in the denominator. Denominator is represented as Denominator in the Sample Calculation listed at the end of this document. Letter d equals 8 patients in the sample calculation.

6. Start Numerator

7. Check Patient Prescribed Inhaled Bronchodilator Therapy AND Results of FEV1<60% Predicted and Patient has COPD Symptoms:
a. If Patient Prescribed Inhaled Bronchodilator Therapy AND Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Met.
b. Reporting Met and Performance Met letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter a equals 4 patients in Sample Calculation.
c. If Patient Prescribed Inhaled Bronchodilator Therapy AND Results of FEV1 <60% Predicted and Patient has COPD symptoms equals No, proceed to check Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator Therapy AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms.

8. Check Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms:
a. If Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.
b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b1 equals 1 patient in the Sample Calculation.
c. If Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals No, proceed to check Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms.

9. Check Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms:
a. If Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.
b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b2 equals 0 patients in the Sample Calculation.
c. If Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals No, proceed to check Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms.

10. Check Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms:
a. If Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.
b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b3 equals 0 patients in the Sample Calculation.
c. If Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals No, proceed to check Spirometry Results FEV1 = 60% Predicted OR Does not have COPD Symptoms.

11. Check Spirometry Results FEV1 = 60% Predicted OR does not have COPD Symptoms:
a. If Spirometry Results FEV1 = 60% Predicted OR Does not have COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.
b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b4 equals 0 patients in the Sample Calculation.
c. If Spirometry Results FEV1 = 60% Predicted OR Does not have COPD symptoms equals NO, proceed to check Spirometry Test Not Performed to Documented, Reason not Given.

12. Check Spirometry Test Not Performed to Documented, Reason Not Given:
a. If Spirometry Test Not Performed to Documented, Reason Not Given equals Yes, include in reporting met and performance exclusion.
b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b5 equals 0 patients in the Sample Calculation.
c. If Spirometry Test Not Performed to Documented, Reason Not Given equals No, proceed to check Inhaled Bronchodilator not Prescribed, Reason Not Specified AND results of FEV1 = 60% Predicted and Patient has COPD Symptoms.

13. Check Inhaled Bronchodilator not Prescribed, Reason Not Specified AND Results of FEV1 = 60% Predicted and Patient has COPD Symptoms:
a. If Inhaled Bronchodilator not Prescribed, Reason not Otherwise Specified AND results of FEV1 = 60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Not Met.
b. Reporting Met and Performance Not Met letter is represented in the Reporting Rate in the Sample Calculation listed at the end of this document. Letter c equals 2 patients in the Sample Calculation.
c. If Inhaled Bronchodilator not Prescribed, Reason not Otherwise Specified AND results of FEV1 = 60% Predicted and Patient has COPD Symptoms equals No, proceed to check Reporting Not Met.

14. Check Reporting Not Met

a. If Reporting Not Met equals No, Quality Data Code or equivalent not reported. 1 patient has been subtracted from reporting numerator in the sample calculation.

Please see Measure Flow in Appendix A.1 for ´Sample Calculation´ referenced above.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure does not require sampling or a survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not Applicable

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Team

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
This is not a composite measure

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0102_MeasureTesting_MSF5.0_Data-635313575706846894.doc,0102_testing_attachment_2015_amended_122915.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 102
Measure Title: COPD: Inhaled bronchodilator therapy
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator

exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

NOTE: As requested, test responses from the 2012 comprehensive review testing form submitted by the AMA-PCPI are included in black font. Due to differences in form structure, sections and questions, and due to different testing, verbatim responses are copied where determined to be most appropriate. What appear to be errors in alignment of responses to questions and/or omissions are likely and are due to these differences. If necessary for clarification, please refer to original 2012 testing form.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Refer to the validity section for a description of the data sample for our EHR testing project.

EHR Measure Validity
The measure was calculated using data collected using two different methods of collection:

- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

The data source was electronic health records in an ambulatory care setting.
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 106 patient encounters.
Data collected from patients seen between 01/01/2010-10/30/2011.
Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

Face Validity

An expert panel was used to assess face validity of the measure. This panel consisted of 12 members, with representation from a number of specialties, including internal medicine, methodology, pulmonology, family medicine, critical care medicine, emergency medicine, pharmacy science, nursing, and health plan representation.

Co-Chairs:
William E. Golden, MD, FACP (University of Arkansas College of Medicine)
Linus Santo Tomas, MD, MS (American College of Chest Physicians)

Members:
Bruce Bagley, MD (American Academy of Family Physicians)
Troy T. Fiesinger, MD (American Academy of Family Physicians)
David G. Jaimovich, MD (Society of Critical Care Medicine)
Bruce Krieger, MD (American Thoracic Society)
Thomas W. Lukens, MD, PhD, FACEP (American College of Emergency Physicians)
Deborah Patterson, MS, RN (Blue Cross Blue Shield Association)
Sam J. W. Romeo, MD, MBA (Tower Health & Wellness Center)
Ralph M. Schapira, MD (VA Medical Center)
Sean D. Sullivan, RPh, PhD (Department of Pharmacy, University of Washington)
Dennis E. Richling, MD (Midwest Business Group on Health)

2015 submission

The data source is the Centers for Medicare & Medicaid Services (CMS) Medicare Physician Quality Reporting System (PQRS) Group Practice Reporting Option (GPRO) Web Interface.

The testing was conducted by Mathematica Policy Research as a component of the 2012 Quality and Resource Use Report (QRUQ), part of the CMS Physician Feedback Reporting Program.

Citation:

Mathematica Policy Research. Experience Report for the Performance Year 2012 Quality and Resource Use Reports. January 8, 2014. Accessed December 7, 2015. Accessible at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2012-QRUR_Experience_Report.pdf

Note that this measure was also tested in 2012 by the PCPI to support NQF re-endorsement for the 2012 comprehensive review. Those test results are not repeated here, however, are available as an attachment on the NQF submission form.

1.3. What are the dates of the data used in testing? January 2012 – December 2012

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice

☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☒ other: ACO (Pioneer and MSSP) with >25 EPs

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Testing and analysis included 66 PQRS GPRO groups with at least 25 eligible professionals (EPs) and 396 ACOs. Groups were included if they reported at least 20 eligible cases for the measure. The groups were distributed across all states, the District of Columbia, Guam and Puerto Rico.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Testing and analysis included 11,593,241 Medicare beneficiaries identified on claims associated with the groups described in 1.5. Beneficiaries attributed to groups with more than 25 EPs averaged 2,974 (standard deviation = 5,105). Approximately half (52%) of the groups were attributed fewer than 1,000 beneficiaries. Beneficiaries attributed to groups with more than 100 EPs averaged 7,077 (standard deviation = 7,842).

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The data were used for reliability testing only. Face validity testing was done with a survey. Other analyses were not done or not applicable.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patients in the testing and analysis were Medicare beneficiaries. No other sociodemographic variables were available for analysis.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Refer to the validity section for a description of the analytic methods for our EHR testing project.

2015 submission

The method of reliability testing as used by Mathematica Policy Research is described as:

"For each of these measures, reliability was estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error):

"Reliability = Variation between groups/(Variation between groups + Variation within group)

"If a score is deemed highly reliable, we would expect that a group's performance rates would be very similar if performance were calculated on the basis of a random sample of the practice's beneficiaries.

"Reliability scores are represented on a continuum from zero to one. Scores closer to zero indicate lower reliability and scores closer to one indicate higher reliability. Although there is no universally agreed-upon minimum reliability threshold, reliability scores in the 0.40–0.70 range are often considered moderate, and scores greater than 0.70 are considered high."

Please see 1.2 for citation.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Refer to the validity section for the testing results for our EHR testing project.

2015 submission

As noted above, scores above 0.70 are considered high.
The reliability for this measure among groups with 25 or more EPs participating in the PQRS GPRO program was 0.85.
The reliability for this measure among ACOs was not included in the analysis.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

We believe this measure remains reliable based on high reliability test scores and relatively large test sample size.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:
• Percent agreement at the denominator and numerator (exception - for those measures with exception)
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

Face Validity
Face validity of the measure score as an indicator of quality was systematically assessed as follows.
After the measure was fully specified, the expert panel (workgroup membership) was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Scale 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5= Strongly Agree

2015 submission

Face validity testing was conducted on the updated measure language. Face validity of the measure score as an indicator of quality was systematically assessed using the following approach:

After the measure was fully specified, the ATS Clinical Practice Committee was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

The rating scale used was 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

The members of the ATS COPD Clinical Practice Committee were selected to serve as an expert panel:

Robert DeMarco, MD
Scott Manaker, MD
Michael Donahoe, MD
Omar Hussain, MD
Katina Nicolacakis, MD
Steve G. Peters, MD
Stephen Hoffman, MD
Alan Plummer, MD
Mike Nelson, MD

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
This measure demonstrates slight agreement when comparing the automated EHR report to visual inspection.

Reliability: N, % Agreement, Kappa
Numerator: 106, 82.08%, kappa 0.1444* (0.0000-0.3248 CI)
Denominator: 106, 100%, kappa non-calculable** (non-calculable CI)

*Refer to the Data Collection Strategy regarding measure implementation changes made at the site.
**Kappa statistic could not be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

Face Validity
The results of the expert panel rating of the validity statement were as follows: N = 7; Mean rating = 4.57 and 100% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality

Frequency Distribution of Ratings
1 - 0 (Strongly Disagree)
2 - 0
3 - 0 (Neither Agree nor Disagree)
4 - 3
5 - 4 (Strongly Agree)

2015 submission

The results of the expert panel rating of the validity statement include:
N = 9
Mean rating = 3.9
Panelists that agree or strongly agree that this measure can accurately distinguish good and poor quality = 88.9%

Frequency distribution of ratings

1 - Strongly disagree	0
2	0
3 - Neither Agree nor Disagree	1
4	8
5 - Strongly Agree	0

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

We believe this measure remains valid based on the degree of agreement by a panel of testers.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 106 patient encounters.
Data collected from patients seen between 01/01/2010-10/30/2011.
Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

Exceptions included medical, patient and system reasons. Exceptions were analyzed for frequency and variability across providers.

2015 submission

Exclusion analysis was not conducted on this measure in this study.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
Although specifications allowed for documented exceptions, including medical, system and patient reasons, for the COPD: Bronchodilator Therapy measure, there were no documented exceptions in this project. All sampled patients were able to be assessed.

2015 submission

Not available

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not available

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

CMS Physician Quality Reporting Initiative:
80,785 cases were reported on for the 2008 program, the most recent year for which data is available.

The following information is for the 2009 program, the only year for which such data is available.
Clinical Condition and Measure: #52 Bronchodilator Therapy
Eligible Professionals: 212,885
Professionals Reporting: 1,336
% Professionals Reporting: 0.63%
# Professionals Reporting >=80% of eligible instances: 424

% Professionals Reporting >=80% of eligible instances: 31.74%

CMS Physician Quality Reporting Initiative/System:
The inter-quartile range (IQR) was calculated to determine the variability of performance on the measure.

2015 submission

This analysis was not conducted on this measure in this study.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Scores on this measure: N = 80,785; Mean = 46.39%
10th percentile: 0.00%
25th percentile: 1.75%
50th percentile: 30.77%
75th percentile: 74.07%
90th percentile: 89.44%

The inter-quartile range (IQR) provides a measure of the dispersion of performance. The IQR is 72.32 and indicates that 50% of physicians have performance on this measure ranging from 1.75% and 74.07% and 10% of physicians have performance rates less than or equal to 0%. (1)

(1)Confidential CMS PQRI 2008 Performance Information by Measure. Jan-Sept TAP file.

2015 submission

Not available

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Not available

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record*

abstraction for the numerator). ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

The measure was calculated using data collected using two different methods of collection:
- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

Data from a performance report for the measure automatically-generated from the EHR (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) were compared to data elements found and scores calculated manually on visual inspection of the medical record by trained abstractors.

Data analysis included:
• Percent agreement at the denominator and numerator(exception - for those measures with exception)
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

2015 submission

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity
This measure demonstrates slight agreement when comparing the automated EHR report to visual inspection.

Reliability: N, % Agreement, Kappa
Numerator: 106, 82.08%, kappa 0.1444* (0.0000-0.3248 CI)
Denominator: 106, 100%, kappa non-calculable** (non-calculable CI)

*Refer to the Data Collection Strategy regarding measure implementation changes made at the site.
**Kappa statistic could not be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

2015 submission

Not applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Missing data analysis was not conducted on this measure in this study.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Not available

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not available

3. Feasibility
Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order). 3a.1. **Data Elements Generated as Byproduct of Care Processes.** generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition If other:
3b. Electronic Sources The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified. 3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)* ALL data elements are in defined fields in electronic claims 3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** 3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-**

specific URL.
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
We have received comments that implementation of this measure remains complex. We agree the complexity of this measure, particularly from eligible professionals without programmable EHRs, could limit it´s use. Several documentation templates are available to facilitate data capture. ATS plans to assess ways to increase awareness of these templates to facilitate reporting.

_____
Statement from 2012 comprehensive review submitted by PCPI to provide history.

The agreement rate and kappa statistic are not indications of a measure problem rather a measure implementation problem. During the first six months of the data collection period, the site did not collect the "documentation of presence of symptoms" in a discrete or searchable field. These findings were generally documented in the history and physical note. After an assessment of its collection methods, the site created a documentation template to capture the components of the numerator.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
N/A

RESPONSE TO ITEM 3b.2
ATS is considering eMeasures in the future. At this time we have not determined whether this measure will be converted.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Quality Improvement (Internal to the specific organization)	Public Reporting Physician Compare https://www.medicare.gov/physiciancompare/staticpages/data/aboutthedata.html

	Payment Program CMS PQRS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

This measure is planned for integration into the CMS Physician Compare Program. Although Physician Compare has been launched, this measure is not yet included as of 12/14/15. The purpose of the Physician Compare Program is to help Medicare beneficiaries make informed choices about health care. The program is broadly available through the Physician Compare website noted above.

This measure has been in use for the CMS PQRI/S program since 2007. The PQRS is a quality reporting program to encourage individual eligible professionals and group practices to report quality information to Medicare. Effective in 2015, the PQRS will be used to apply a negative payment adjustment to individual eligible professionals and group practices who do not satisfactorily report data on quality measures for covered professional services provided to Medicare patients in 2013.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Please see 1b.2 for performance results, including favorable trends from 2011-2014. Please see 1b.4 for performance on disparities. No further analyses have been conducted on improvement.

Findings from the National Ambulatory Medical Care survey 1999-2010 show progressive increase in use of bronchodilators over time (Ford, et al., 2014). However, the study is limited in the ability to determine whether or not this is an appropriate increase per guideline recommendations.

Citation:

Ford ES, Mannino DM, Giles WH, Wheaton AG, Liu Y, Croft JB. Prescription practices for chronic obstructive pulmonary disease: findings from the national ambulatory medical care survey 1999-2010. COPD. 2014 Jun;11(3):247-55.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
This measure was initially developed for use in the PQRI/S program. Although not publically reported until the recent release of the new Physician Compare site, the performance trends indicate progress on increasing use of inhaled bronchodilators for the

management of COPD from 2011-2014.

The ATS supports the goal of high-quality, efficient healthcare. Toward that goal, the ATS Quality Improvement Committee reviews performance annually as a component of measure maintenance and plans further analyses in the future. ATS participates in international COPD guideline development as well as conducts educational sessions and an annual meeting featuring use of guidelines, including appropriate bronchodilator use.

As noted in 1b.2, we have updated the specifications to this measure in an effort to improve appropriate long-acting inhaled bronchodilator use and further improve COPD patient care quality.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

We are not aware of any unintended consequences related to this measurement.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses 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 Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

COMMENT ON 5a.1 - N/A is not a selection. For this reason, we select yes. There are no competing measures to harmonize.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 0102_measure_flow_2015_form_no_s.19_.doc

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Thoracic Society
Co.2 **Point of Contact:** Gary, Ewart, gewart@thoracic.org, 202-296-9770-
Co.3 **Measure Developer if different from Measure Steward:** Northfield Associates LLC
Co.4 **Point of Contact:** Sue, Frechette, sue.frechette@northfieldassoc.com, 802-496-7815-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
This measure was initially developed in 2007 by the AMA-PCPI, working with the ATS.

PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

The initial Work Group Panel consisted of:
William E. Golden, MD, FACP, co-chair
Linus Santo Tomas, MD, MS, co-chair
Bruce Bagley, MD (AAFP)
Troy T. Fiesinger, MD (AAFP)
David G. Jaimovich, MD (SCCM)
Bruce Krieger, MD (ATS)
Thomas W. Lukens, MD, PhD, FACEP (ACEP)
Susan Nedza, MD, MBA, FACEP (CMS)
Deborah Patterson, MS, RN (BCBSA)
Sam J. W. Romeo, MD, MBA
Ralph M Schapira, MD (VA)
Sean D. Sullivan, RPh, PhD
Dennis E. Richling, MD
Nancy Lawler, RN (Joint Commission)

Stewardship of this measure was transferred to the ATS in November 2014.

To prepare for the 2015 NQF comprehensive review, ATS formed the Quality Improvement Committee Sub-committee on COPD Measures to review and update this measure. The Sub-committee members include:

Laura Feemster, MD, MS, VA Puget Sound Health Care System, University of Washington Medical Center, Chair
Bela Patel, MD, The University of Texas Health Science Center at Houston
Carolyn Fruci, MD, Prima-CARE, PC
David Au, MD, MS, VA Puget Sound Health Care System, University of Washington Medical Center
Jerry A. Krishnan, MD, PhD, University of Illinois at Chicago
Gary Ewart, Chief, Advocacy & Government Relations, ATS
Sue Frechette, RN, MBA, Consultant, Northfield Associates LLC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2007
Ad.3 **Month and Year of most recent revision:** 12, 2011
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.6 **Copyright statement:** The Measures are not clinical guidelines, do not establish a standard of medical care, and have not been tested for all potential applications.

The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measures require a license agreement between the user and the PCPI® Foundation (PCPI®) or the American Thoracic Society (ATS). Neither ATS, nor the American Medical Association (AMA), nor the AMA-convened Physician Consortium for Performance Improvement® (AMA-PCPI), now known as PCPI, nor their members shall be responsible for any use of the Measures.

The AMA's and AMA-PCPI's significant past efforts and contributions to the development and updating of the Measures is acknowledged. ATS is solely responsible for the review and enhancement ("Maintenance") of the Measures as of September 8, 2014.

ATS encourages use of the Measures by other health care professionals, where appropriate.

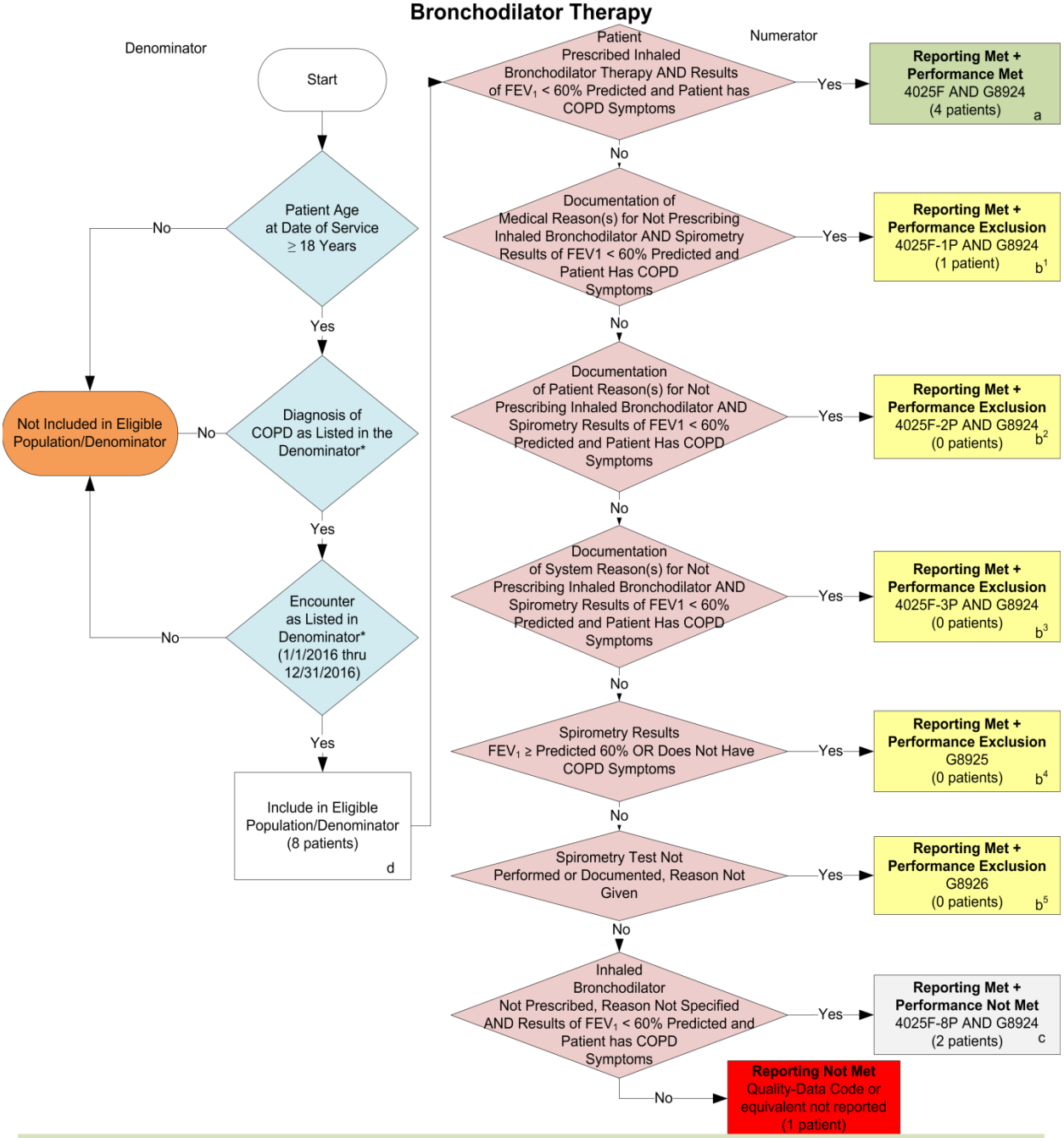
THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. ATS, the AMA, the PCPI and its members and former members of the AMA-PCPI disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** Coding/Specifications updates occur annually. ATS plans to continue measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. ATS will also review the measures if there is a major change in scientific evidence, results from testing or other issues are noted that materially affect the integrity of the measure.

2016 Claims Individual Measure Flow
PQRS #52 NQF #0102: Chronic Obstructive Pulmonary Disease (COPD): Inhaled Bronchodilator Therapy



SAMPLE CALCULATIONS:

Reporting Rate=

Performance Met (a=4 patients) + Performance Exclusion ($b^1+b^2+b^3+b^4+b^5$=1 patient) + Performance Not Met (c=2 patients) / Eligible Population / Denominator (d=8 patients) = 7 patients / 8 patients = **87.50%**

Performance Rate=

Performance Met (a=4 patients) / Reporting Numerator (7 patients) – Performance Exclusion ($b^1+b^2+b^3+b^4+b^5$= 1 patient) = 4 patients / 6 patients = **66.67%**

*See the posted Measure Specifications for specific coding and instructions to report this measure.
NOTE: Reporting Frequency – Patient-process

2016 Registry Individual Measure Flow
PQRS #52 NQF #0102: Chronic Obstructive Pulmonary Disease (COPD): Inhaled Bronchodilator Therapy

SAMPLE CALCULATIONS:

Reporting Rate=

Performance Met (a=4 patients) + Performance Exclusion ($b^1+b^2+b^3$=1 patient) / Eligible Population / Denominator (d=8 patients) = 7 patients / 8 patients = **87.50%**

Performance Rate=

Performance Met (a=4 patients) / Reporting Numerator (7 patients) – Performance Exclusion ($b^1+b^2+b^3$= 1 patient) = 4 patients / 6 patients = **66.67%**

*See the posted Measure Specification for specific coding and instructions to report this measure.
NOTE" Reporting Frequency – Patient-process

v2

2016 Claims Individual Measure Flow
PQRS #52 NQF #0102: Chronic Obstructive Pulmonary Disease (COPD): Inhaled Bronchodilator Therapy

Please refer to the specific section of the Measure Specification to identify the denominator and numerator information for use in reporting this Individual Measure.

1. Start with Denominator
2. Check Patient Age:
    a. If the Age is greater than or equal to 18 years of age on Date of Service and equals No during the measurement period, do not include in Eligible Patient Population. Stop Processing.
    b. If the Age is greater than or equal to 18 years of age on Date of Service and equals Yes during the measurement period, proceed to check Patient Diagnosis.
3. Check Patient Diagnosis:
    a. If Diagnosis of COPD as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Diagnosis of COPD as Listed in the Denominator equals Yes, proceed to check Encounter Performed.
4. Check Encounter Performed:
    a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Encounter as Listed in the Denominator equals Yes, include in the Eligible population.
5. Denominator Population:
    a. Denominator population is all Eligible Patients in the denominator. Denominator is represented as Denominator in the Sample Calculation listed at the end of this document. Letter d equals 8 patients in the sample calculation.
6. Start Numerator
7. Check Patient Prescribed Inhaled Bronchodilator Therapy AND Results of FEV1<60% Predicted and Patient has COPD Symptoms:
    a. If Patient Prescribed Inhaled Bronchodilator Therapy AND Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Met.
    b. Reporting Met and Performance Met letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter a equals 4 patients in Sample Calculation.
    c. If Patient Prescribed Inhaled Bronchodilator Therapy AND Results of FEV1 <60% Predicted and Patient has COPD symptoms equals No, proceed to check Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator Therapy AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms.
8. Check Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms:
    a. If Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.
    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b1 equals 1 patient in the Sample Calculation.
    c. If Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals No, proceed to check Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms.

9. Check Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms:

    a. If Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b2 equals 0 patients in the Sample Calculation.

    c. If Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals No, proceed to check Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms.

10. Check Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms:

    a. If Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b3 equals 0 patients in the Sample Calculation.

    c. If Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator AND Spirometry Results of FEV1 <60% Predicted and Patient has COPD Symptoms equals No, proceed to check Spirometry Results FEV1 ≥ 60% Predicted OR Does not have COPD Symptoms.

11. Check Spirometry Results FEV1 ≥ 60% Predicted OR does not have COPD Symptoms:

    a. If Spirometry Results FEV1 ≥ 60% Predicted OR Does not have COPD Symptoms equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b4 equals 0 patients in the Sample Calculation.

    c. If Spirometry Results FEV1 ≥ 60% Predicted OR Does not have COPD symptoms equals NO, proceed to check Spirometry Test Not Performed to Documented, Reason not Given.

12. Check Spirometry Test Not Performed to Documented, Reason Not Given:

    a. If Spirometry Test Not Performed to Documented, Reason Not Given equals Yes, include in reporting met and performance exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b5 equals 0 patients in the Sample Calculation.

    c. If Spirometry Test Not Performed to Documented, Reason Not Given equals No, proceed to check Inhaled Bronchodilator not Prescribed, Reason Not Specified AND results of FEV1 ≥ 60% Predicted and Patient has COPD Symptoms.

13. Check Inhaled Bronchodilator not Prescribed, Reason Not Specified AND Results of FEV1 ≥ 60% Predicted and Patient has COPD Symptoms:

    a. If Inhaled Bronchodilator not Prescribed, Reason not Otherwise Specified AND results of FEV1 ≥ 60% Predicted and Patient has COPD Symptoms equals Yes, include in Reporting Met and Performance Not Met.

    b. Reporting Met and Performance Not Met letter is represented in the Reporting Rate in the Sample Calculation listed at the end of this document. Letter c equals 2 patients in the Sample Calculation.

    c. If Inhaled Bronchodilator not Prescribed, Reason not Otherwise Specified AND results of FEV1 ≥ 60% Predicted and Patient has COPD Symptoms equals No, proceed to check Reporting Not Met.

14. Check Reporting Not Met

    a. If Reporting Not Met equals No, Quality Data Code or equivalent not reported. 1 patient has been subtracted from reporting numerator in the sample calculation.

SAMPLE CALCULATIONS:

Reporting Rate=

$$\frac{\text{Performance Met (a=4 patients) + Performance Exclusion } (b^1+b^2+b^3+b^4+b^5=1 \text{ patient}) + \text{Performance Not Met (c=2 patients)}}{\text{Eligible Population / Denominator (d=8 patients)}} = \frac{7 \text{ patients}}{8 \text{ patients}} = \mathbf{87.50\%}$$

Performance Rate=

$$\frac{\text{Performance Met (a=4 patients)}}{\text{Reporting Numerator (7 patients)} - \text{Performance Exclusion } (b^1+b^2+b^3+b^4+b^5= 1 \text{ patient})} = \frac{4 \text{ patients}}{6 \text{ patients}} = \mathbf{66.67\%}$$

2016 Registry Individual Measure Flow

PQRS #52 NQF #0102: Chronic Obstructive Pulmonary Disease (COPD): Inhaled Bronchodilator Therapy

Please refer to the specific section of the Measure Specification to identify the denominator and numerator information for use in reporting this Individual Measure.

1. Start with Denominator
2. Check Patient Age:
    a. If the Age is greater than or equal to 18 years of age on Date of Service and equals No during the measurement period, do not include in Eligible Patient Population. Stop Processing.
    b. If the Age is greater than or equal to 18 years of age on Date of Service and equals Yes during the measurement period, proceed to check Patient Diagnosis.
3. Check Patient Diagnosis:
    a. If Diagnosis of COPD as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Diagnosis of COPD as Listed in the Denominator equals Yes, proceed to check Encounter Performed.
4. Check Encounter Performed:
    a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Encounter as Listed in the Denominator equals Yes, proceed to check
5. Check Spirometry test results demonstrate FEV1 < 60% Predicted and patient has COPD symptoms:
    a. If Spirometry test results demonstrate FEV1 < 60% Predicted and patient has COPD symptoms equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Encounter as Listed in the Denominator equals Yes, include in the Eligible population.
6. Denominator Population:
    a. Denominator population is all Eligible Patients in the denominator. Denominator is represented as Denominator in the Sample Calculation listed at the end of this document. Letter d equals 8 patients in the sample calculation.
7. Start Numerator

8. Check Patient Prescribed Inhaled Bronchodilator Therapy:

    a. If Patient Prescribed Inhaled Bronchodilator Therapy equals Yes, include in Reporting Met and Performance Met.

    b. Reporting Met and Performance Met letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter a equals 4 patients in Sample Calculation.

    c. If Patient Prescribed Inhaled Bronchodilator Therapy equals No, proceed to check Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator Therapy.

9. Check Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator:

    a. If Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b1 equals 1 patient in the Sample Calculation.

    c. If Documentation of Medical Reason(s) for Not Prescribing Inhaled Bronchodilator equals No, proceed to check Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator.

10. Check Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator:

    a. If Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b2 equals 0 patients in the Sample Calculation.

    c. If Documentation of Patient Reason(s) for Not Prescribing Inhaled Bronchodilator equals No, proceed to check Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator.

11. Check Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator:

    a. If Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator equals Yes, include in Reporting Met and Performance Exclusion.

    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b3 equals 0 patients in the Sample Calculation.

    c. If Documentation of System Reason(s) for Not Prescribing Inhaled Bronchodilator equals No, proceed to check Inhaled Bronchodilator not Prescribe, Reason Not Specified.

12. Check Inhaled Bronchodilator not Prescribe, Reason Not Specified:

    a. If Inhaled Bronchodilator not Prescribe, Reason not Specified equals Yes, include in Reporting Met and Performance Not Met.

    b. Reporting Met and Performance Not Met letter is represented in the Reporting Rate in the Sample Calculation listed at the end of this document. Letter c equals 2 patients in the Sample Calculation.

    c. If Inhaled Bronchodilator not Prescribe, Reason not Specified equals No, proceed to check Reporting Not Met.

13. Check Reporting Not Met

    a. If Reporting Not Met equals No, Quality Data Code or equivalent not reported. 1 patient has been subtracted from reporting numerator in the sample calculation.

SAMPLE CALCULATIONS:

Reporting Rate=

$$\frac{\text{Performance Met (a=4 patients) + Performance Exclusion } (b^1+b^2+b^3=1 \text{ patient}) + \text{Performance Not Met (c=2 patients)}}{\text{Eligible Population / Denominator (d=8 patients)}} = \frac{7 \text{ patients}}{8 \text{ patients}} = \mathbf{87.50\%}$$

Performance Rate=

$$\frac{\text{Performance Met (a=4 patients)}}{\text{Reporting Numerator (7 patients)} - \text{Performance Exclusion } (b^1+b^2+b^3=1 \text{ patient})} = \frac{4 \text{ patients}}{6 \text{ patients}} = \mathbf{66.67\%}$$

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0275
Measure Title: Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)
Measure Steward: Agency for Healthcare Research and Quality
Brief Description of Measure: Admissions with a principal diagnosis of chronic obstructive pulmonary disease (COPD) or asthma per 1,000 population, ages 40 years and older. Excludes obstetric admissions and transfers from other institutions.

[NOTE: The software provides the rate per population. However, common practice reports the measure as per 100,000 population. The user must multiply the rate obtained from the software by 100,000 to report admissions per 100,000 population.]
Developer Rationale: This indicator is intended to identify hospitalizations for chronic obstructive pulmonary disease (COPD) in older adults age 40 years and older. With access to high quality care and community resources that promote prevention and self-care for emphysema, chronic bronchitis, and asthma, rates of COPD and exacerbations requiring hospitalization may be reduced.

Numerator Statement: Discharges, for patients ages 40 years and older, with either
• a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for COPD (excluding acute bronchitis); or
• a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for asthma

[NOTE: By definition, discharges with a principal diagnosis of COPD or asthma are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QI software does not explicitly exclude obstetric cases.]
Denominator Statement: Population ages 40 years and older in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.
Denominator Exclusions: n/a

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Population : County or City

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported

by the stated rationale.

Summary of evidence:

- This is a population-level measure. The level of analysis is county or city.
- New evidence is provided since the last endorsement maintenance review.
- A systematic review of the body of evidence is not required for outcome measures
  - The developer provides the following rationale: Access to high quality care and community resources that promote improved population health, combined with appropriate self-care for emphysema, chronic bronchitis, and asthma, may reduce the rates of COPD and exacerbations requiring hospitalizations.
- Although not required per NQF guidance, the developer conducted a literature review (January 2012-October 2015) related to aspects of hospitalization for asthma, as follows:
  - Geographic and temporal variation
  - Disparities
  - Environmental exposure
  - Access to care

Questions for the Committee:

- *Although the developer provides updated evidence related to access to care for COPD, does the Committee agree the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on Evidence?*
- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer reports:

- COPD is one of the most common chronic diseases in the United States, and is currently the third leading cause of death.
- In 2011, COPD was the 7th most common non-newborn principal diagnosis for hospital admissions. The table below shows that in the AHRQ QI 2013 PQI Reference Population there were 655,570 qualifying hospital discharges for COPD, with a population rate of 4.4 per 1,000.

Reference Population Rate and Distribution of County Performance for PQI 05
Overall Reference Population Rate

Year	Number of Counties	Number of Events (Numerator)	Population at Risk (Denominator)	Observed Rate Per 1,000
2009	3,137	808,770	140,912,452	5.7395
2010	3,141	785,156	143,145,444	5.485
2011	3,143	802,121	145,239,701	5.5227
2012	3,141	759,704	147,030,305	5.167
2013	3,140	655,570	148,812,929	4.4053

Distribution of County-level Observed Rates in Reference Population Per 1,000
(p=percentile)

Year	Number of Counties	Mean	SD	p5	p25	Median	p75	p95
2009	3,137	7.10	25.61	0.04	3.33	5.73	8.50	15.16
2010	3,138	7.10	31.11	0.02	3.22	5.44	8.11	13.87

2011	3,141	6.78	25.55	0.08	3.15	5.36	8.09	13.81
2012	3,139	6.67	31.06	0.08	2.89	4.98	7.67	12.96
2013	3,140	5.12	4.11	0.00	2.49	4.52	6.92	11.93

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Disparities

The developer notes:

- Counties with higher incomes have lower COPD hospitalization rates than counties with lower incomes.
- Blacks had approximately 50% higher hospitalization rates than Whites and nearly double the hospitalization rates of other ethnic groups.
- In 2010, there was little variation between men and women (33.4 vs 31.6 hospital visits per 10,000 civilian population for women and men, respectively)
- In 2009 to 2010, there was little variation between blacks and whites (34.9 vs 30.5 hospital visits per 10,000 civilian population for Blacks and Whites, respectively)
- Age-adjusted rates were highest among Native American and lowest among Asian in most years.

Question for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**Yes. The underlying rationale for the measure remains reasonable and new evidence provided since the last endorsement maintainance review supports that.

**Outcomes measure ties directly to the evidence. The desired outcome would be a reduction in admissions for COPD or Asthma in Older Adults. No concerns with the evidence to support the measure.

**no need for repeat discussion; rationale provides optoins for providers to achieve a change in measure results

**NB: 73% of the variability of COPD hospitalization relative risk was attributed to unidentified regional social and physical environments shared by HSAs rather than to unique local HAS factors.

I think we should re-discuss given above.

Not sure that the provider can do anything to change measure results.

**Evidence and rationale seems strong for such a measure as there are many external factors that might influence the rate of admissions for COPD/ASTHMA in adults including the environment, access to care, quality of care by providers in the associated community, and these would be expected to be fairly direct but some more modifiable than others.

**Measure is a health outcome. Rationale supports relationship between home health care and community resources and reduced rates of hospitalization for COPD among persons 40+ years old.

**This is an outcome measure.

Hospital admission for COPD/asthma for those >40 per population in county over 40.

The stated data for an intervention to reduce this rate appears to be inconsistent.

How does this measure provide more actionable information than other population-based measures (e.g. smoking rates)?

1b. Performance Gap

Comments:

**Yes. Updated data continues to show this measure is warranted. Counties with higherincomes have lower COPD hospitilization rates than counties with lower incomes. The updated data provided shows a gap in care that warrants a national performance measure.

**Performance data on the measure was provided. It did demonstrate a gap in care or suboptimal performance. It further demonstrated that lower socio-economic status was a factor in the outcomes (admissions). It further identified that there was a gap in performance between for black Americans.

**performance gap in care exists that warrants a national performance measure

**The measure highlights gaps in care and highlights the potential need for more targeted interventions for more disparate groups
**Moderate.
Not sure how to reconcile to bullet points below:

- Blacks had approximately 50% higher hospitalization rates than Whites and nearly double the hospitalization rates of other ethnic groups.
- In 2009 to 2010, there was little variation between blacks and whites (34.9 vs 30.5 hospital visits per 10,000 civilian population for Blacks and Whites, respectively)

**Performance data does suggest a gap.
The evidence for disparities is a bit noisy; while there is some evidence for disparities especially by race or income, other referenced data suggest that at lest for some years there was no difference between black and whites (e.g 2009-2010).
**Performance gaps were demonstrated across US counties, adequate to warrant a national performance measure. Subgroup differences indicate higher rates in older adults 65+ yrs old, women and minority groups, as well as higher in rural areas. The disparities exist and might be amenable to prevention and improved self-care.
**There are provided variations in median income, rural vs urban, gender, and region. Race and ethnicity were not examine d- perhaps because missing in 16% of records?

Priority is reported as 1c.
The stewards describe as "high priority" due to frequency, morbidity/mortality, resource use and consequences of poor care. I find these accurate although question how much is poor care vs SES, environmental, etc issues.

1c. High Priority (previously referred to as High Impact)
Comments:
**No concerns.
**N/a
**NA
**n/a
**N/A. Though it includes 3 diagnostic entities, they are not used as subscores.
**Not applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Administrative claims

Specifications:

- The developer has made changes to the measure specifications since the last endorsement review.
- The numerator for this measure is: *the number of discharges, for patients age 40 years and older, with either a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for COPD (excluding acute bronchitis); or a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for asthma*
- The denominator is: *the population ages 40 years and older in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.*
- The calculation algorithm is stated in S.18.
- This outcome measure is risk adjusted, using a statistical risk model.

Questions for the Committee :

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*

2a2. Reliability Testing [Testing attachment]
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Summary of reliability testing is not available from the prior review.

Describe any updates to testing

- The developer indicates there are updates to the reliability testing since the last submission
  - Reliability testing at the level of the measure score has been conducted using more current data.
  - Data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) is used
    - 40 states representing about 89% of U.S. county hospital discharges, for a total of about 30 million hospital discharges from community hospitals
    - The HCUP dataset included data from 2009-2013
  - The developer created a county characteristic dataset, merged with county-level observed and risk adjusted rates from the HCUP dataset. Candidate predictor variables from the County Health Rankings (CHR) dataset and the American Community Survey (ACS) were analyzed.
    - The CHR dataset included data from 2014.
    - The ACS dataset included data from 2013.
- The developer provides two risk models:
  - age and gender composition of the county; and
  - optional addition to these two variables of the percent of households falling below the federal poverty level

- Validity analyses: Sociodemographic variables were combined using principal component analysis (PCA) into a single socioeconomic status (SES) variable, defined at the county level.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Reliability testing was conducted at the performance measure level, but not the individual element level.
- Testing was conducted using signal-to-noise analysis assessing the reliability to confidently distinguish the performance among counties.

Results of reliability testing

- The developer reported a signal-to-noise ratio of 0.97, which the developer states indicates strong reliability. The developer states reliability is strong for all county sizes.
- The developer reported that when SES is added to the risk adjustment, the signal-to-noise ratio is 0.96.

Guidance from the Reliability Algorithm : 1 → 2 → 4 → 5 → 6 (highest eligible ratings is HIGH).

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Summary of validity testing is not available from the prior review.

Describe any updates to validity testing

- Updates have been made to the validity testing since the last submission
- Empirical validity testing of the measure score was conducted.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- The measure was tested for face validity with input from four clinical expert panels for a total of 73 panelists. The panel was convened from 2008-2009.
- Empirical validity testing also was conducted by correlating the measure score to various factors, including health behaviors, access to care, etc.
- Based on the RAND-UCLA Appropriateness Method, the panel was a "nominal group" panel.

Validity testing results:

The developer reports:

- For empirical testing, the developer reports:
  - Health behaviors (HB) were only statistically significant predictors ($p<.00010$) (see Table 3a).
  - Access to care (AC) was not significant when HB and SES/E are included in the model.
  - Categorizing the variables into interpretable groups (HB, AC, and SES/E) resulted in significant collinearity in the models. In particular, there was a correlation of magnitude 0.77 between IHB and one of the components for SES/E and correlations of magnitude between 0.40 and 0.50 between the components for AC and the other two components for SES/E (see Table 3b). Hence the relative importance of those factors should be interpreted with caution.
- The developer further notes that its disparities analysis found zip codes in the highest income quartile have 31% lower admission rates than those in the lowest income quartile. The developer states that, from a population health perspective, such disparities argue for the importance of the indicator in capturing poor outcomes for vulnerable populations and that by taking a population health perspective, efforts to decrease individual risk factors, such as obesity, smoking, and limited physical exercise (the variables included in the health behaviors factor) may decrease prevalence and exacerbations of asthma.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer reports the following:

- Excludes cases with missing gender, age, quarter, year, principal diagnosis or county. Per the developer, these exclusions never exceed 1% of eligible records.
- 682 discharges were excluded due to diagnoses of cystic fibrosis and anomalies of the respiratory system. Removing this exclusion would increase the numerator count by 0.10%. The exclusion of cystic fibrosis and anomalies of the respiratory system has been retained to increase the face validity of the measure and to align the measure with the young adult and pediatric PQI/PDI for asthma admissions. Patients are identified without additional burden.
- Patients with severe chronic respiratory diseases have been excluded because COPD/asthma in the context of these diseases differs clinically from the patients with COPD/asthma alone.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

The developer reports:

- The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect)
  - The covariates are gender and age (in 5-year age groups)
  - An option model is available that includes percent of households under the federal poverty level
  - A conceptual model acknowledging the impact of community factors also was considered (e.g., clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise, community norms and beliefs, etc., which can impact hospitalization rates).
- The risk adjustment model was calibrated by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.
  - The developer reports the observed to predicated values across the deciles range between 0.95-1.03, indicating it is well calibrated.
  - The developer reports the c-statistic (measure of well the risk adjustment model distinguishes events from non-events, is the c-statistic) was 0.53, which it concludes was low and presumes was due to the limited predictors included.
  - The developer noted the addition of SES to the model improves the calibration (range of 0.89-1.08), but that the c-statistic did not change.

Questions for the Committee:

- *Is the risk adjustment methodology appropriate?*
- *Were the appropriate community factors included in the conceptual model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer assessed the probability that a county is higher or lower than a benchmark or threshold, given county size—i.e., whether the indicator can discriminate the best performing counties from the lower performing counties. The developer reports on discrimination results for counties:

- Measure has strong discrimination to:
  - identify low performing counties for most counties: 85% of counties can be classified as better or worse

than the threshold
    - identify high performing counties for moderate to large counties; 85% of counties can be classified as better or worse than the benchmark.
- Performance discrimination remains strong when adding SES to risk adjustment.

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

The developer notes:

- The AHRQ QIs use frequently reported administrative data variables. PQI 05 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis.
- Missing data handled – rates are typically less than 1% of the state database.

Guidance from the Validity Algorithm : 1 → 2 → 3 → 6 → 7 → 8 (highest eligible rating is HIGH)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Specifications are consistent with the evidence.
**The measure description and data indicates strong validity.
**specifications are consistent with evidence
**None
**yes
**Validity was tested by face validity and by testing. Testing included comparisons with a number of data sources. The results of the validity testing suggests that some factors are important in impacting the measure such as Health Behaviors, whereas others are not such as access to care or SES. There was evidence that high economic zip codes had 31% lower admission rates. While there seems to be sufficient evidence that the ,measure may reflect those actors that travel with income and also with health behaviors, the validity does not seem so strong for factors related to intensity or quality of medical care in the associated community.
**I did not review the ICD codes used for this, but it is likely appropriate and can be validated by a clinical expert. I have identified no issues. Risk adjustment and structure of the measure are appropriate.
**Prior data demonstrated face validity, with "some concern". These concerns are noted, but may be obviated in the use in a public health setting.
**This appears relatively straightforward. There is the potential challenge of changes in residence with possibilities of errors in attribution. I suspect this is a minor threat.
How might differences in coding practice affect the measured rate? For example, do hospitals in areas with high SES have more resources to ensure COPD patients on NPPV are coded as acute on chronic resp failure as primary diagnosis – excluding them from this measure?
How is the confidence in the O/E presented? Considering different population in different counties, I can imagine less precision based on county population. How is this presented? I note in 2b5.2 that in deciles of counties with smaller populations, more are "unclassified" – presumably due to this lack of precision. But how is this imprecision presented by those using the measure?
**Validity testing included empirical testing and systematic assessment of face validity.

2a2. Reliability Testing

Comments:

**The measure was tested for face validity with imut from four clinical expert panels for a total of 73 panelists. Empirical validity testing was also conducted by correlating the measure score with various factors including health behaviors, access to care, etc. Disparities analysis found zip codes in the highest income quintile have 31% lower admission rates than those in the lowest income quintile.
**Validity testing was conducted. Based on the methodology, the measure as specified should be an indicator of quality. Better outpatient care should result in lower admissions.
**sufficient validity so that conclusions about quality can be made
**Adequate
**High. Yes, sufficient validity.
Concerned by statement that "access to care was not significant when HB and SES are included in the model"
**As above; I am skeptical that the validity testing supports the measure being sensitive to variations in quality of care.

**Reliability testing using HCUP data up to 2013 seems well done and acceptable, done at the county level. Good reliability was demonstrated.
**Reliability was assessed by signal-to-noise ratio. The lowest (0.84) was found in the 314 counties with the smallest average population. It is unclear to me if counties with very small populations are excluded as I expect the signal-to-noise to worse with smaller and smaller populations.

2b2. Validity Testing
Comments:
**Excluded cases never exceed 1% of eligible records and they are consistant wit the evidence.The risk adjustment methodology including appropriate community factors appears appropriate.Performance discrimination remainss strong when adding SES to risk adjustment as this measure identifies meaningful differences in quality.
**The missing data did not constitute a threat to the validity of the measure. The remaining data was adequate to draw results.
**exclusions are c/w evidence, no groups inappropriately excluded; risk adjustment methodology appropriate; measure identifies meaningful differences in quality
**Missing data are excluded -- no concerns
**Face validity was performed in 2008-2009. I am unsure if the same results would be found if performed today. The panels noted that observation status may affect this indicator. With the change in the two-midnight rule, I am unclear how much of the decrease in admission rates reported is real vs due to this impact. This causes me to question the validity (as well as the reliability over time).
There is description of the panel rating a level of support for each measure. I do not see that reported for this measure.
Empirical validity found that health behaviors was the only predictor in a negative binomial model. HB was collinear with SES and access to care, however. The stewards argue that "taking a population health perspective, efforts to decrease individual risk factors…may decerase prevalence and exacerbations of COPD." Why then aren't these the measures (which are more directly actionable) than this one?
**2b3. Yes, No
**Recent empiric validity testing is sufficient. Although access to care is not significantly related to the measure, the strong mediation by health behaviors and environment/SES indicate the utility of the measure for public health purposes. I believe the measure could be used as an indicator of quality, albeit mediated by health behaviors, environment and SES. The health systems and public health should be working on all the levels, but especially on health behaviors to impact this measure.
**There are minimal exclusions based on patient diagnoses.
2b4. Risk assessment seems appropriate. Appropriate community factors are included in the model. Not clear then if the proposal is to risk adjust this measure. Is there a signal for quality if risk-adjusted?
**The discrimination of the model (as measured by c statistic) is 0.53 – largely the same as a coin flip. That means, that there is a 53% chance that a randomly selected case will have a risk score higher than a randomly selected control. This suggests the predictors are largely independent of the outcome.
2b5. Yes, meaningful differences do seem to be present given strong discrimination.
**I cannot determine if this measure identifies meaningful differences in quality or meaningful differences in the population. This section notes "strong discrimination" but the prior section suggests this is not the case (at least based on AUC).
2b6. NA
**This is listed as N/A. Multiple risk adjusting models are mentioned but not clearly articulated, compared or a recommendation among them made.
2b7. No issue
**I think the biggest threat to validity is based on the relative contributions to the measure from structural and background factors, patient behaviors and provider behaviors. These result may reflect the quality of the overall asthma/COPD experience, but less well reflect the quality of provider behavior
Risk adjustment was not robust using limited factors and resulting in a low c score
Meaningful differences in quality of the overall disease control may be present but the part attributable to quality of care is less clear.
Missing data - It could but it does not appear to be a quantitatively significant issue.
**Unclear – it is reported that missing data rates are "typically less than 1%" but these data are not specifically supplied.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments: **Reliability testing at the level of the measure score has been conducted using more current data from the Healthcare Cost and Utilization Project State Inpatient Databases. The developer has also included two risk models and Validity analyses defined

at the county level. The net result is strong reliability for all county sizes.
**Reliability testing was performed with an adequate population size. The measure has sufficient reliability.
**sufficient reliability demonstrated
**Adequate
**6, High
**The reliability testing seems to support a high level of reliability. It would be interesting to understand why the SD has suddenly plummeted during the last years of data reported (31 to 4; 2012 to 2013).
**None of these issues indicate a threat to validity to me.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer reports:

- All data elements are in defined fields in electronic claims.
- The measure is based on readily available administrative billing and claims data and U.S. Census data.
- There are no fees. The AHRQ QI software is publically available at no cost.
- Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**All data elements are in defined fields in electronic claims, the measure is based on readily available administrative billing and claims data and US census data, the AHRQ QI software is publically available at no cost and users have over 10 years of experience using the AHRQ QI software in SAS and Windows.
**The measure is feasible and the data is administrative data. It should, over time, be an electronic record measure.
**all data elements are routinely generation and used during care delivery
**No concerns
**High
**Feasibility is high as the data are all available from electronic sources.
**This is highly feasible. The required elements are available in electronic form.
**Appears feasible by definitions provided. It is not clear to me how far in arrears the data are available.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
The measure is currently being used in the following programs:
Payment programs:

- CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use

Reports (QRUR)
- Oregon Health Authority
- CMS Medicare Shared Savings Program

Regulatory and Accreditation Program:
- Statewide Quality Advisory Committee (Massachusetts)

Quality Improvement:
- West Jefferson Medical Center

Public Reporting:
- Arizona Department of Health Services, AZ Hospital Compare
- CMS Medicaid Adult Core Measures1
- Connecticut Department of Health Services, CT Hospital Compare
- Maine Health Data Organization (MHDO),
- Nevada Compare Care
- Oklahoma State Department of Health, MONAHRQ
- Utah Department of Health, MONAHRQ website
- Virginia Health Information, MONAHRQ website
- Washington State, MONAHRQ website
- California Office of Statewide Health Planning and Development, Healthcare Information Division
- Connecticut, Office of Health Care Access
- St John´s Episcopal Hospital
- Arkansas Department of Human Services: Arkansas Medicaid Performance
- Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
- Northwest Hospital and Medical Center

Improvement results

The developer reports the following:
- The PQI 05 hospital admissions rate has decreased by 104,000 fewer hospitalizations from 2011 to 2013
  - In 2012 the rate was 5.2 per 1,000
  - in 2013 the rate was 4.4 per 1,000
- Variation among counties decreased "substantially" in 2013. The developer is not certain whether this is a single year anomaly or an actual trend.

Unexpected findings (positive or negative) during implementation
- The developer reports no challenges implementing this measure. It also notes the use of the AHRQ PQIs has grown since the initial endorsement, suggesting the measure is highly implementable and useful.
- The transition to ICD-10-CM will provide challenges in understanding time trends or rates from calendar year 2015. This is an expected challenge all measures based on coded data will encounter.

Potential harms

The developer did not identify potential harms.

Feedback: No feedback provided on QPS. Measure reviewed by MAP for Physician Quality Reporting System (PQRS) and the Medicare Shared Saving Program in 2013. The measure was also reviewed in Value-Based Payment Modifier (VBPM) Program in 2012. MAP 2013 decision to support the measure in the PQRS. MAP 2012 decision do not support the measure in the VBPM program.

***Questions for the Committee**:*
- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement
4c. Unintended Consequences
Comments:
**The measure is currently being used in numerous Payment programs, Massachusetts Statewide Quakity Advisory Committee, Quality improvement at West Jefferson Medial Center and numerous Public Reporting State Health Services.
**The results of this measure indicate quality outcomes and perhaps quality of life. Improving the rates should lead to high-quality, efficient healthcare with reduced admissions.
**performance results can be used to further the goal of high-quality, efficient health care.
**No unintended consequences, no concerns with use and usability
**Yes, seems usable given the # of groups using the measure. Improvement seen from 2011 to 2013.
**The measure is being used fairly widely, including by over fifteen entities including CMS and multiple health agencies, and several hospitals, .There do not appear to be any problems with implementation or unintended consequences. The measure does seem to be showing improvement over time but as I stated above it is unclear to me as to how much of this is attributable to improvements in provider quality of care vs improvements in health behaviors, or even environmental factors including changes in smoking regulations and reduction of second hand smoke
**The measure is widely used at local, state and national level. It can drive public health activities, behavioral health, and health system approaches to QI, and demonstrated reductions indicate major cost savings have been achieved in recent years as COPD admissions have declined. Outpatient care is much more economical and reduction in high/outlier counties has driven widespread savings.
There are no identifiable unintended consequences.
**The stewards mention a number of entities using the measure.
They also describe improvements. It was not clear to me if the improvements were only in those areas using the measure or across the US. Why did these improvement occur? Was it because of a rise in observation status hospitalizations? How might this affect reliability of the measure over time? How might disparities to access to such status affect county-level data?

Criterion 5: Related and Competing Measures
Related or competing measures • No related or competing measures identified.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0275

Measure Title: Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Hospital admissions for Chronic Obstructive Pulmonary Disease (COPD) or asthma in older adults

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

This indicator is intended to identify hospitalizations for obstructive lung conditions that progress to a severity requiring inpatient care in older adults (age 40 years and older). There is often significant clinical diagnostic overlap between the chronic lung diseases including emphysema, chronic bronchitis, and asthma. This overlap and diagnostic uncertainty is greater in older adults. Medical therapies, home-based clinical interventions, comprehensive care and pulmonary rehabilitation are associated with lower hospitalization rates in some, but not all studies. The COPD hospitalization rate is associated with pollution levels and smoking rates, suggesting community level leverage points. Access to high quality care and community resources that promote improved population health (e.g., healthy behaviors, risk factor reduction and prevention), when combined with appropriate self-care for emphysema, chronic bronchitis, and asthma, may reduce the rates of rates of COPD and exacerbations requiring hospitalization.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Please note that this is an outcome measure, so a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.8 below, to provide additional context and support for the measure.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

Not applicable

1a.4.1. Guideline citation (*including date*) and URL for guideline (*if available online*):

Not applicable

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Not applicable

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches are performed to continually update evidence for all AHRQ PQIs, including PQI 05. The current evidence review results presented below constitute articles published from January 2012 - October 2015. Additional articles from previous years were identified from prior project-related literature reviews. Search terms included the relevant MeSH term (pulmonary disease, chronic obstructive). We combined this clinical search string with (hospital*[Title/Abstract]) AND (prevent*[Title/Abstract] OR "access to care"[Title/Abstract] OR "ambulatory care sensitive"[Title/Abstract] OR “avoidable hospitalization”[Title/Abstract] OR "small area analysis"[MeSH]). For completeness we also tested more inclusive search strings. Below we have provided a summary of the most up-to-date evidence.

1a.8.2. Provide the citation and summary for each piece of evidence.

Chronic obstructive pulmonary disease (COPD) consists of three primary diseases: asthma, emphysema, and chronic bronchitis. Though each disease causes respiratory dysfunction, each has somewhat distinct etiologies, treatments, and outcomes. Since admission for asthma specifically (in those <40) is considered in a separate indicator, asthma will not be discussed in this section. Only the evidence for COPD as it relates to emphysema and chronic bronchitis will be discussed.

Importance

COPD is one of the most common chronic diseases in the U.S., and is currently the third leading cause of death[1]. In 2011, 6.5% of adults reported having a COPD diagnosis.[1,2] In 2011, COPD was the 7th most common non-newborn principal diagnosis for hospital admissions.[3] Table 1 (see Measure Testing form) shows that in the AHRQ QI 2013 PQI Reference Population there were 655,570 qualifying hospital discharges for COPD with a population rate of 4.4 per 1,000.[4] The coefficient of variation was 0.86 in 2013.[4]

Geographic and Temporal variation

A study of Medicare data by Ford et al. (2013) reported that changes in age-adjusted rates from 1999 to 2010 varied between states.[2] A comparison of state-specific Medicare hospital rates in 1999-2000 to those in 2009-2010 demonstrates geographic clustering of the 10 states in 1999 to 2000, with the highest hospitalization rates (14.0-26.6 per 1,000 Medicare enrollees) along the Mississippi River and Ohio River valleys. By 2009 to 2010, there was a marked improvement in rates in many of those states. States with the highest age-adjusted Medicare hospitalization rates in 2009 to 2010 are similar to those states with the highest age- adjusted prevalence of COPD in 2011. During 1999 to 2010 no state experienced a significant increase in age-adjusted Medicare hospitalization rates while 26 states experienced significant declines ($P < 0.05$).

This same study also reported temporal changes in COPD hospitalization during this timeframe. Using NHDS data, the study reported declining trends for age-adjusted rates for COPD hospitalization during 1999 to 2010 among all adults ($P = 0.001$), men ($P < 0.001$), and women ($P = 0.022$). Similarly, using Medicare data, the study found age-adjusted rates for COPD hospitalizations declined during 1999 to 2010 for men ($P = 0.022$) and for all enrollees overall ($P = 0.045$), but the decline was not significant for women or specific race groups.

Another large national study of Medicare data by Holt et al. (2011) likewise reported inter and intrastate variation and geographic clustering of COPD hospitalization rates between 1995 and 2006.[5] According to the study, health service area level (HSA-level) COPD hospitalization rates had a median of 11.7 and a range of 3.0 to 76.3. With few exceptions, states with the highest and lowest ranges of intrastate rates were highly clustered geographically. Excessive hospitalization risk was concentrated in Appalachia, the southern Great Lakes, the Mississippi Delta, the Deep South, and west Texas. In the Bayesian spatial mixture model, 73% of variability of COPD hospitalization relative risk was attributed to unidentified regional social and physical environments shared by HSAs rather than to unique local HSA factors (27%).

Validity

As a population health indicator, we consider the relationship of PQI 05 to a wide range of factors that are amenable to changes in public policy and community based interventions and may, in turn, reduce the aggregate hospitalization rates or lower the incidence of COPD.

Disparities

Healthy People 2020 defines inequities/disparities in outcomes as "A particular type of health difference that is closely linked with social or economic disadvantage. Health disparities adversely affect groups of people who have systematically experienced greater social or economic obstacles to health based on their racial or ethnic group, religion, socioeconomic status, gender, mental health, cognitive, sensory, or physical disability, sexual orientation, geographic location, or other characteristics historically linked to discrimination or exclusion." Disparities in outcomes among subpopulations highlight the need for improvement in vulnerable populations. Counties with higher incomes consistently have lower COPD hospitalization rates than counties with lower incomes (See Table 1 of the Measure Testing Form). Race disparities are less clear. California's Office of Statewide Health Planning and Development found that Blacks consistently had about 50% higher hospitalization rates than Whites and nearly double the hospitalization rates of other ethnic groups.[6] However, one retrospective study of National Hospital Discharge Survey (NHDS) and Medicare data by Ford et al. (2013) found that age-adjusted rates of hospitalizations for COPD varied little between men and women in 2010 (33.4 vs 31.6 hospital visits per 10,000 civilian population for women and men, respectively) or between Blacks and Whites during 2009 to 2010 (34.9 vs 30.5 hospital visits per 10,000 civilian population for Blacks and Whites, respectively).
It should be noted, however, that the percent of hospital records missing race information ranged from 16.0% to 31.0% for the NHDS.[2] Based on Medicare data, the study found that age-adjusted rates were highest among Native American enrollees and lowest among Asian enrollees in most years (statistical significance not provided).

The following section discusses evidence related to leverage points within the community outside the healthcare system. It describes the relationship of county characteristics to COPD hospitalization. Effective interventions to modify these characteristics may impact COPD hospitalization, although all studies only evaluated the relationship and not causality.

Environmental exposure

One important aspect of population health is the impact of environmental exposure on the health of residents in a community. Poor air quality can increase the likelihood of acute exacerbations resulting COPD-related hospitalizations. A systematic review and meta-analysis by Song et al. (2014) investigating the quantitative effects of outdoor air pollution (represented by a 10 μg/m3 increment of particulate matter [PM]10) reported that the acute effect of exposure and the total effect estimate on COPD hospital admissions was 1.02 (95% CI: 1.01–1.02).[7] Short-term exposure to 10 ug/m3 increment of PM10 led to COPD hospital admission increases of 1% in China (95% CI 1.01-1.01), 2% in US (95% CI 1.01-1.03) and 1% in EU (95% CI 1.00-1.02); the heterogeneity across studies was non-significant. Another recent study by Yap et al. (2015) examined the impact that an ordinance aimed at reducing residential wood burning had on particulate matter levels and hospitalizations in California's San Joaquin Valley Air Basin (SJVAB).[8] While it observed statistically significant reductions in particulate matter and coarse particles after the implementation of the ordinance ($p < 0.05$) and a decrease in COPD hospital admission rates per 1000 for adults from 7.2 to 6.5, the findings for hospital admission rates were not statistically significant.

Access to care

The following section discusses evidence related to leverage points within the health care system. The following studies primarily examine alternative care models and patient-level hospitalization rates. While it cannot be assumed from these studies that improving care for patients will necessarily result in lower area level COPD hospitalization rates, these studies demonstrate potential mechanisms to improve outcomes for patients.

Several studies have shown that home-based clinical interventions have lowered hospitalization rates. In one clinical trial, team-managed home-based primary care decreased hospital readmissions by about 22% in patients with severe disability at 6 months ($p = .03$).[9] Two other studies showed that in-home healthcare for patients with advanced COPD reduced the cost of care, hospitalizations, and visits to the emergency department.[10,11 12,13] Also, a home monitoring service for patients with COPD reduced hospitalization rates in the UK.[14]
Several additional studies have examined comprehensive care and self-care assistance programs. In a multicenter randomized trial designed to assess the effect of self-management interventions, hospital admissions for exacerbation of COPD were reduced by 39.8% ($p = .01$) and ED visits by 41.0% ($p = .02$) in the intervention group compared with the usual care group.[15] A second study found that patients receiving individually tailored care plans, and access to web- and phone-based nurse case managers experienced lower hospital readmission rates within one year of discharge for COPD exacerbation ($p = .03$).[16] Another Dutch study of a web-based interactive COPD symptom surveillance and educational program linked to telephone-based nurse specialists found that participants experienced a significantly reduced rate of inpatient admissions ($p = .02$) and emergency unit visits ($p = .03$).[17] A cohort of COPD patients in New York and Connecticut graduating from a minimum 10-session pulmonary rehabilitation program across 11 outpatient, hospital-based centers experienced fewer respiratory-specific hospitalizations in the year following program completion compared to the year prior to rehabilitation (0.18 vs. 0.35 admissions per person per year, $p = .008$).[18] In contrast, one randomized trial examined short-term and long-term effects of an outpatient pulmonary rehabilitation program for patients with COPD. The intervention group experienced a significant ($p < .0001$) reduction in exacerbations, but not reductions in the number of hospitalizations.[19] Two studies focusing on educational efforts among Veterans Affairs patients failed to demonstrate a statistically significant impact on COPD hospitalizations: one multi-site randomized trial by Fan et al. (2012) seeking to determine the efficacy of a comprehensive care management program (CCMP) found no significant differences in COPD hospitalization rates between the intervention and control groups when the study was terminated early due to serious safety concerns,[20] and an educational and self-efficacy intervention assessed by Siddique et al. (2012) likewise reported no statistically significant impact on COPD hospitalizations.[21]

Influenza has been shown to be associated with COPD hospital admissions, suggesting prevention of influenza in vulnerable populations may impact hospitalization rates. Using time-series regression models, Gerke et al. (2013) found a strong, significant association between concurrent influenza activity and COPD hospitalizations ($p < 0.0001$).[22] The association was especially strong among older patients (attributable risk measure* for COPD patients <65 vs >=65 yrs is 0.0374 and 0.0419, respectively). However, a study by Seo et al. (2013) using multivariate logistic analysis showed that influenza vaccination did not significantly reduce the risk of hospitalization due to new onset or acute exacerbation of chronic pulmonary disease among Korean patients.[23]

Smoking increases risk of respiratory symptoms and COPD.[24] Smoking cessation is associated with lower risk of hospitalization. An Australian study[25] found, amongst study groups, the current smokers group was over six times more likely to be hospitalized for COPD compared with never smokers group (hazard ratio of 6.81 (95% CI 5.87–7.89)). The excess risk of hospitalization and risk advancement period (RAP) for COPD was reduced within 5 years of smoking cessation across all age groups. A RAP of 17.7 years was found among current smokers compared to never smokers for COPD. Among those who had smoked for 25 years or longer, the risk of hospitalization for COPD was significantly higher in those who smoked more than 15 cigarettes per day, compared to those who smoked less than 15 cigarettes per day.

References

1. Hoyert DL, Xu J. Deaths: preliminary data for 2011. Natl Vital Stat Rep. 2012;61(6):1-52.
2. Ford ES, Croft JB, Mannino DM, Wheaton AG, Zhang X, Giles WH. COPD surveillance--United States, 1999-2011. Chest. 2013;144(1):284-305.
3. Pfuntner A, Wier LM, Stocks C. Most frequent conditions in US hospitals, 2011. 2013.
4. Agency for Healthcare Research and Quality R, MD. HCUP State Inpatient Databases (SID). HCUP Cost and Utilization Project (HCUP): . 2013;AHRQ QI Software Version 6.0.
5. Holt JB, Zhang X, Presley-Cantrell L, Croft JB. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. International journal of chronic obstructive pulmonary disease. 2011;6:321-328.
6. (OSPHD) COoSHPaD. Racial and Ethnic Disparities in Healthcare in California: California Fact Book 2010. Winter 2010.
7. Song Q, Christiani DC, XiaorongWang, Ren J. The global contribution of outdoor air pollution to the incidence, prevalence, mortality and hospital admission for chronic obstructive pulmonary disease: a systematic review and meta-analysis. International journal of environmental research and public health. 2014;11(11):11822-11832.
8. Yap PS, Garcia C. Effectiveness of residential wood-burning regulation on decreasing particulate matter levels and hospitalizations in the San Joaquin Valley Air Basin. American journal of public health. 2015;105(4):772-778.
9. Hughes SL, Weaver FM, Giobbie-Hurder A, et al. Effectiveness of team-managed home-based primary care: a randomized multicenter trial. Jama. 2000;284(22):2877-2885.
10. Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. European Respiratory Journal. 2003;21(1):58-67.
11. Steinel JA, Madigan EA. Resource utilization in home health chronic obstructive pulmonary disease management. Outcomes management. 2002;7(1):23-27.
12. Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. Eur Respir J. 2003;21(1):58-67.
13. Steinel JA, Madigan EA. Resource utilization in home health chronic obstructive pulmonary disease management. Outcomes Manag. 2003;7(1):23-27.
14. Dale J, Connor S, Tolley K. An evaluation of the west Surrey telemedicine monitoring project. J Telemed Telecare. 2003;9 Suppl 1:S39-41.
15. Bourbeau J, Julien M, Maltais F, et al. Reduction of hospital utilization in patients with chronic obstructive pulmonary disease: a disease-specific self-management intervention. Arch Intern Med. 2003;163(5):585-591.
16. Casas A, Troosters T, Garcia-Aymerich J, et al. Integrated care prevents hospitalisations for exacerbations in COPD patients. Eur Respir J. 2006;28(1):123-130.
17. Trappenburg JC, Niesink A, de Weert-van Oene GH, et al. Effects of telemonitoring in patients with chronic obstructive pulmonary disease. Telemed J E Health. 2008;14(2):138-146.
18. Raskin J, Spiegler P, McCusker C, et al. The effect of pulmonary rehabilitation on healthcare utilization in chronic obstructive pulmonary disease: The Northeast Pulmonary Rehabilitation Consortium. J Cardiopulm Rehabil. 2006;26(4):231-236.
19. Guell R, Casan P, Belda J, et al. Long-term effects of outpatient rehabilitation of COPD: A randomized trial. Chest. 2000;117(4):976-983.
20. Fan VS, Gaziano JM, Lew R, et al. A comprehensive care management program to prevent chronic obstructive pulmonary disease hospitalizations: a randomized, controlled trial. Annals of internal medicine. 2012;156(10):673-683.
21. Siddique HH, Olson RH, Parenti CM, et al. Randomized trial of pragmatic education for low-risk COPD patients: impact on hospitalizations and emergency department visits. International journal of chronic obstructive pulmonary disease. 2012;7:719-728.
22. Gerke AK, Tang F, Yang M, Foster ED, Cavanaugh JE, Polgreen PM. Predicting chronic obstructive pulmonary disease hospitalizations based on concurrent influenza activity. Copd. 2013;10(5):573-580.
23. Seo YB, Hong KW, Kim IS, et al. Effectiveness of the influenza vaccine at preventing hospitalization due to acute lower respiratory infection and exacerbation of chronic cardiopulmonary disease in Korea during 2010-2011. Vaccine. 2013;31(10):1426-1430.

24. Pleasants RA, Heidari K, Wheaton AG, et al. Targeting Persons With or At High Risk for Chronic Obstructive Pulmonary Disease by State-based Surveillance. Copd. 2015:1-10.
25. Tran B, Falster MO, Douglas K, Blyth F, Jorm LR. Smoking and potentially preventable hospitalisation: the benefit of smoking cessation in older ages. Drug and alcohol dependence. 2015;150:85-91.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PQI05_NQF_0275_Evidence_Form_151214.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This indicator is intended to identify hospitalizations for chronic obstructive pulmonary disease (COPD) in older adults age 40 years and older. With access to high quality care and community resources that promote prevention and self-care for emphysema, chronic bronchitis, and asthma, rates of COPD and exacerbations requiring hospitalization may be reduced.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental files.
Table 1. Reference Population Rate and Distribution of County Performance for PQI 05
Overall Reference Population Rate

Year	Number of Counties	Number of Events (Numerator)a	Population at Risk (Denominator)a	Observed Rate Per 1,000a
2009	3,137	808,770	140,912,452	5.7395
2010	3,141	785,156	143,145,444	5.485
2011	3,143	802,121	145,239,701	5.5227
2012	3,141	759,704	147,030,305	5.167
2013	3,140	655,570	148,812,929	4.4053

Distribution of County-level Observed Rates in Reference Population Per 1,000

Year	Number of Counties	(p=percentile)b Mean	SD	p5	p25	Median	p75	p95
2009	3,137	7.10	25.61	0.04	3.33	5.73	8.50	15.16
2010	3,141	7.10	31.11	0.02	3.22	5.44	8.11	13.87
2011	3,143	6.78	25.55	0.08	3.15	5.36	8.09	13.81
2012	3,141	6.67	31.06	0.08	2.89	4.98	7.67	12.96
2013	3,140	5.12	4.11	0.00	2.49	4.52	6.92	11.93

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare

Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)
aThe observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible (ages 40 and above) population of all counties included in the reference population data (denominator). Note: Observations from counties with rates outside of 1.5*interquartile range are excluded as outliers.
bThe distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile. Note: Counties with rates outside of 1.5*interquartile range are excluded as outliers.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
n/a

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental files.
Table 2. Admission Rates per 1,000 (PQI 05), by patient and hospital characteristics, 2013

Patient/hospital characteristic	Estimate	Std Error	p-value (Ref Grp = *)	Lower 95% CL	Upper 95% CL
Total U.S.	440.50	0.5423		439.44	441.56
Patient Characteristics					
Age Groups:					
18-39	.	.	.	.	
40-64*	337.21	0.7325		335.77	338.64
65 and over	566.64	0.8077	<.001	565.05	568.22
Gender:					
Male*	387.99	0.8045		386.41	389.56
Female	484.28	0.7342	<.001	482.84	485.72
Patient Zip Code Median Income					
First quartile (lowest income)	636.04	1.9296	<.001	632.26	639.83
Second quartile	564.24	1.2921	<.001	561.70	566.77
Third quartile	456.48	1.1177	<.001	454.29	458.67
Fourth quartile (highest income)*	360.02	0.7598		358.53	361.51
Location of patient residence (NCHS):					
Rural	518.53	3.8991	<.001	510.88	526.17
Urban*	438.95	0.5476		437.88	440.02
Location of Care:					
Northeast*	412.248	1.249		409.80	414.70
Midwest	533.233	1.157	<.001	530.96	535.50
South	481.255	0.890	<.001	479.51	483.00
West	315.038	1.162	<.001	312.76	317.32

Source: Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project, Nationwide Inpatient Sample, 2013, and AHRQ Quality Indicators, version 6.0.
Rates are adjusted by age and gender using the AHRQ QI PQI Reference Population for 2013 as the standard population; when reporting is by age, the adjustment is by gender only; when reporting is by gender, the adjustment is by age only.
NCHS - National Center for Health Statistics designation for urban-rural locations.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
n/a

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
COPD is one of the most common chronic diseases in the U.S. and is the third leading cause of death1. In 2011, 6.5% of adults reported having a COPD diagnosis(2). In 2011, COPD was the 7th most common non-newborn principal diagnosis for hospital admissions(3). Table 1 shows that in the AHRQ QI 2013 PQI Reference Population there were 655,570 qualifying discharges with a rate of 4.4 per 1,000(4). The coefficient of variation was 0.80 in 2013(4).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Hoyert DL, Xu JQ. Deaths: preliminary data for 2011. Natl Vital Stat Rep. 2012;61(6):1-65. Hyattsville, MD: National Center for Health Statistics.2012.
2. Ford ES, Croft JB, Mannino DM, Wheaton AG, Zhang X, Giles WH. COPD surveillance--United States, 1999-2011. Chest. 2013;144(1):284-305.
3. Pfuntner A, Wier LM, Stocks C. Most Frequent Conditions in U.S. Hospitals, 2011. HCUP Statistical Brief #162. September 2013. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb148.pdf. Accessed August 19, 2013.
4. HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
n/a

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Prevention, Pulmonary/Critical Care : Asthma, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://1.usa.gov/1JTj6wq Note: The URL link currently provides Version 5.0 specifications. Version 6.0 specifications will be released publicly March 2016 and found via the module page: http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PQI05_Technical_Specifications_v6.0_151214v02.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
As standard protocol, the AHRQ QI program annually updates all measures with Fiscal Year coding changes, refinements based on stakeholder input, refinements to improve specificity and sensitivity based on additional analyses, and necessary software changes. In addition, approximately every two years, AHRQ updates the risk adjustment parameter estimates and composite weights based on the most recent year of data (i.e., the most current reference population possible). The refined measures are tested and confirmed to be valid and reliable prior to release of the updated software.

Since the last update, the following changes have been made to the indicator:
- Added exclusion for discharges with any-listed ICD-9-CM diagnosis codes for cystic fibrosis and anomalies of the respiratory system.
- Added inclusion criterion for discharges with a principal ICD-9-CM diagnosis code for asthma. This was added after a clinical panel review of the indicator based on the diagnostic uncertainty between asthma and COPD in older adults
- Changed eligible age range for numerator and denominator to age 40 and above following clinical panel review. Asthma in younger adults (18-39 years) is captured in PQI 15.
- Removed inclusion criterion for discharges with a principal ICD-9-CM diagnosis code for acute bronchitis and any secondary ICD-9-CM diagnosis codes for COPD (excluding acute bronchitis). This inclusion criterion is no longer necessary given revised coding guidelines.
- Removed formal exclusion for discharges in MDC 14. By definition, discharges with a principal diagnosis of COPD, asthma, or acute bronchitis are precluded from an assignment of MDC 14 by grouper software.
- The data upon which to base the reference population was updated.
- Updated with 2013 US Census population estimates
- Fiscal Year coding updates

Additional information regarding revisions to PQI software and technical specifications available online:
http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx
Note: Version 6.0 specifications will be released publicly March 2016.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Discharges, for patients ages 40 years and older, with either
- a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for COPD (excluding acute bronchitis); or
- a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for asthma

[NOTE: By definition, discharges with a principal diagnosis of COPD or asthma are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QI software does not explicitly exclude obstetric cases.]

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Users may specify a time period; but the time period is generally one year. Note that the reference population rates and signal variance parameters assume a one-year time period.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.

Please see attached excel file in S.2b. for Version 6.0 specifications.
Prevention Quality Indicators technical specifications and appendices also available online at http://www.qualityindicators.ahrq.gov/Modules/PQI_TechSpec.aspx). Note: The URL link currently provides Version 5.0 specifications. Version 6.0 specifications will be released publicly March 2016.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Population ages 40 years and older in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The term "metropolitan area" (MA) was adopted by the U.S. Census in 1990 and referred collectively to metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs), and primary metropolitan statistical areas (PMSAs). In addition, "area" could refer to either 1) FIPS county, 2) modified FIPS county, 3) 1999 OMB Metropolitan Statistical Area, or 4) 2003 OMB Metropolitan Statistical Area. Micropolitan Statistical Areas are not used in the QI software.

See AHRQ QI website for 2014 Population File Denominator report for calculation of population estimates embedded within AHRQ QI software programs. http://www.qualityindicators.ahrq.gov/Downloads/Software/SAS/V50/AHRQ_QI_Population_File_V50.pdf

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
n/a

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
n/a

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
n/a

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The predicted value for each case is computed using a logistic regression with covariates for gender and age (in 5-year age groups). An option model is available that includes percent of households under the federal poverty level as well. Because we cannot individually observe the age and gender of each person in a counties population, we use the age and gender distribution of the county to estimate the number of "cases" in each age*gender group. The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Data (SID) for the year 2013 (combined), a database consisting of 40 states and the U.S. Census data by county. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., area). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Additional information on methodology can be found in the Empirical Methods document on the AHRQ Quality Indicator website (www.qualityindicators.ahrq.gov) and attached in the supplemental information.

The specific covariates for this measure are as follows:

PARAMETER	LABEL
SEX	Female
AGE	Male, Age 40-44
AGE	Male, Age 45-49
AGE	Male, Age 50-54
AGE	Male, Age 55-59
AGE	Male, Age 60-64
AGE	Male, Age 65-69
AGE	Male, Age 70-74
AGE	Male, Age 75-79
AGE	Male, Age 80-84
AGE	Male, Age 85+
AGE	Female, Age 40-44
AGE	Female, Age 45-49
AGE	Female, Age 50-54
AGE	Female, Age 55-59
AGE	Female, Age 60-64
AGE	Female, Age 65-69
AGE	Female, Age 70-74
AGE	Female, Age 75-79
AGE	Female, Age 80-84
AGE	Female, Age 85+
POVCAT	Poverty Decile 2
POVCAT	Poverty Decile 3
POVCAT	Poverty Decile 4
POVCAT	Poverty Decile 5
POVCAT	Poverty Decile 6
POVCAT	Poverty Decile 7
POVCAT	Poverty Decile 8
POVCAT	Poverty Decile 9
POVCAT	Poverty Decile 10 (Highest percent poverty)*

*Deciles are based on the percentage of households under the federal poverty level (FPL).
Source: http://qualityindicators.ahrq.gov/Modules/pqi_resources.aspx
Parameter estimates with and without SES covariates (POVCAT) are included with the Technical Specifications.
Please note Version 6.0 will be released publicly in March 2016.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
n/a

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 observed rate of each PQI is simply the number of individuals living in a county admitted to the hospital for the condition of interest divided by the census population estimate for the area (for PQI 05 ages 40 and above). The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected performance observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic distributions observed in the user's dataset? The expected rate is calculated only for risk-adjusted indicators.

The expected rate is estimated for each county using logistic regression.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the performance observed in the user's dataset were applied to a mix of patients with demographics distributed like the reference population. The risk adjusted rate is calculated using the indirect method as observed rate divided by expected rate multiplied by the reference population rate. The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural counties).

For additional information, please see supporting information in the Quality Indicator Empirical Methods attached in the supplemental files.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing), or county (PSTCO=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications and software are specified to be used with any ICD-9-CM- or ICD-10-CM/PCS coded administrative billing/claims/discharge dataset.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Population : County or City

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: all community based care

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
n/a

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PQI05_NQF_0275_Measure_Testing_Form_151214v02.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0275
Measure Title: Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in

understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items.

Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2009-2013. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and

Quality (AHRQ).[1] HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 40 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

The SID data elements include ICD-9-CM coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

For additional testing of the indicators we created a county characteristic dataset which was merged with the county-level observed and risk adjusted rates from the HCUP dataset described above. Using a conceptual model for hospitalization indicators, we examined candidate predictor variables from the County Health Rankings (CHR)[2] dataset and the American Community Survey (ACS)[3]. Candidate variables that corresponded to the conceptual model were grouped by the following categories: individual health behavior (IHB) included variables that include actions and behaviors of individuals and may be mutable, access to care (AC) included variables that reflect the structure and quality of the healthcare system in a community, socioeconomic status (SES) included variables of poverty and education levels within a community and environment (E) variables included community characteristics, such as access to food and open spaces, pollution or violent crime. County prevalence estimates were derived from the Behavioral Risk Factor Surveillance System (BRFSS) COPD model based county estimates.

1.3. What are the dates of the data used in testing?

HCUP data included 2009-2013. The CHR data was from 2014 and ACS data from 2013.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

County

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice

[1] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

[2] "County Health Rankings & Roadmaps". University of Wisconsin Population Health Institute, 2014. http://www.countyhealthrankings.org. Accessed 26 Jan. 2015.

[3] "American Community Survey (ACS)." *https://www.census.gov/programs-surveys/acs/data.html*. Accessed 2015.

☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Population health	☒ other: County

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1. Reference Population Rate and Distribution of County Performance for PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Overall Reference Population Rate				
Year	**Number Counties**	**Outcome of Interest (Numerator)[1]**	**Population at Risk (Denominator)[1]**	**Observed Rate Per 1000[1]**
2009	3,137	808,770	140,912,452	5.7395
2010	3,141	785,156	143,145,444	5.4850
2011	3,143	802,121	145,239,701	5.5227
2012	3,141	759,704	147,030,305	5.1670
2013	3,140	655,570	148,812,929	4.4053

Distribution of County-level Observed Rates in Reference Population Per 1,000								
Year	Number of Counties	(p=percentile)[2]						
		Mean	SD	p5	p25	Median	p75	p95
2009	3,137	7.10	25.61	0.04	3.33	5.73	8.50	15.16
2010	3,141	7.10	31.11	0.02	3.22	5.44	8.11	13.87
2011	3,143	6.78	25.55	0.08	3.15	5.36	8.09	13.81
2012	3,141	6.67	31.06	0.08	2.89	4.98	7.67	12.96
2013	3,140	5.12	4.11	0.00	2.49	4.52	6.92	11.93

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population for the county in the reference population data (denominator). For PQI 05 this includes population age 40 years and older.

[2]The distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

See 1.5 (Table 1)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The reliability and performance discrimination testing was completed using 2013 HCUP data. Validity testing was completed using 2012 HCUP data. The AHRQ QI PQI 2012 reference population has 34,440,38 discharges and the AHRQ QI PQI 2013 reference population includes 29,891,024. Annual testing of rates, reliability and performance discrimination showed little change in performance between the two reference populations.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We use SDS variables for two purposes: risk adjustment and validity analyses. Because the measures are applied at the county level, we use variables that describe the make-up of the county population.

Risk adjustment: Two risk models are available; one includes age and sex make-up of the county, the other also includes the percent of households falling below the federal poverty level. These data are obtained from the US Census.

Validity analyses: In addition to the risk models we used county level demographic variables in testing. Sociodemographic variables were combined using principal component analysis (PCA) into a single socioeconomic status (SES) variable, defined at the county level. Including unemployment rate, percent adults below the federal poverty line, percent of adults aged 25-44 with some post-secondary education, percent of population not proficient in English and median household income.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Signal-to-noise

The signal-to-noise ratio refers to the entire population of US counties, comparing the degree to which rates are different from county to county (the signal) to how stable the rates are within counties (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance between counties, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences between counties (e.g. one county performs better than others).

The signal-to-noise ratio is estimated for each county. The overall signal-to-noise estimate is an average of county-level signal to noise ratios weighted by county size. County size is calculated as the eligible population for PQI 05 (population 40 years and older). Weighting by county size reduces the impact of counties that have very small denominators (the number of patients at risk).

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one county's performance from the other counties in the population, it is sensitive to the distribution of county sizes as well as the distribution of observed rates in the reference population. If the counties in a population all have performance in a narrow range, it is more difficult to reliably distinguish between counties' performance than when county performance is spread out over a much wider range. For example, if all counties have nearly perfect performance, it will be impossible to distinguish between them. As a consequence, if the distribution of county rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal to noise ratio. Different methods of calculating reliability and signal-to-noise result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8. To account for the uncertainty (noise) in a county's performance due to reliability concerns stemming from low volume, smoothed rates can be calculated.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a. Signal-to-Noise Ratio by Size Decile for PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Size Decile	Number of Counties	Avg. Number of Qualifying Population per County in Decile	Avg. Signal-to-Noise Ratio for Counties in Decile
1	314	1553.3	0.84412
2	314	3725.4	0.93023
3	314	5819.1	0.95444
4	314	8390.4	0.96804
5	314	11299.9	0.97562
6	314	15572.9	0.98212
7	314	21779.2	0.98700
8	314	34147.8	0.99153
9	314	65115.8	0.99530
10	314	306522.7	0.99845
Overall	3140	47392.7	0.96693

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 2b. SES Signal-to-Noise Ratio by Size Decile for PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Size Decile	Number of Counties	Avg. Number of Qualifying Population per County in Decile	Avg. Signal-to-Noise Ratio for Counties in Decile
1	314	1553.3	0.80092
2	314	3725.4	0.90464

3	314	5819.1	0.93702
4	314	8390.4	0.95555
5	314	11299.9	0.96600
6	314	15572.9	0.97500
7	314	21779.2	0.98179
8	314	34147.8	0.98810
9	314	65115.8	0.99338
10	314	306522.7	0.99781
Overall	3140	47392.7	0.95699

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

The indicator demonstrates good reliability with a signal-to-noise ratio of 0.97. Reliability remains strong for all county sizes. Smoothed rates, which are recommended for all counties (and are implemented in the AHRQ software), address any remaining reliability concerns for the smallest counties. When SES is added to the risk adjustment, the reliability remains high at 0.96.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Systematic Assessment of Face Validity:

In 2008-2009, we convened clinical panels to assess the use of the PQI across a wide range of applications, one of which was comparative reporting of area level rates.[4] We solicited nominations from national professional organization to form four clinical expert groups for a total of 73 panelists. We utilized a hybrid approach for the panel review, using two review processes, which were conducted simultaneously with information exchange between the two panels and a third and fourth panel that conducted specific reviews based on specialty. The development of this hybrid process builds from the experiences in previous panel evaluations of QI modules. The panel process that has been employed during the development of the PSIs, the PDIs and the validation of the IQIs is based on the RAND-UCLA Appropriateness Method

4 Davies SM, McDonald KM, Schmidt E, Schultz E, Geppert J, Romano, PS. 'Expanding the Uses of AHRQ's Prevention Quality Indicators: Validity from the Clinician Perspective'. Medical Care. 2011 Aut;49(8):479-85.

and is termed a "nominal group" panel. The approach allowed for a wider range of input is fully described elsewhere.[5]

Panelists rated the indicator on appropriateness of use after completing a 14 item questionnaire. The questionnaire evaluated the face validity of the indicators, the panelists' perspectives on bias and potential for gaming, and the overall usefulness of the indicators when applied at one of three aspects of the health care system: area, payer and large provider organizations, for one of three purposes: internal quality improvement, comparative reporting (either public or not), and pay for performance.

Support was defined as follows:

- Full support for use: Median score of 7-9 without disagreement
- Some concern regarding use: Median score of 4-6.9 regardless of agreement status
- General support with some concerns regarding use due to disagreement: Median score of 7-9 with disagreement
- Major concern regarding use: Median score or 1-3.9 regardless of agreement status

Empirical Validity

We sought to assess the relationship of county-level hospital admission rate for COPD with county level measures of socioeconomic status (SES) and community environment, heath behaviors and individual risk factors (i.e. smoking, physical activity and obesity), and access to quality care measures (i.e. primary care physician and other primary care provider density, diabetes screening testing, uninsurance). SES and community environment variables included unemployment rate, poverty rate, some college rate, English proficiency rate, median household income, food environment index, access to exercise, violent crime rate, air pollution, severe housing problems and rural status derived from the Community Health Rankings data and the American Community Survey. Because of the high number of relevant variables, we aimed to improve the interpretability of results. Principal component analysis (PCA) was used to create three composite variables for county characteristics: SES and environment (SES/Environment), health behaviors and personal risk factors (HB) and access to care (AC) using. The variables included in the analyses (described earlier in this paragraph) were determined a priori based on clinical and subject matter expertise. Because the relationship between the individual variables within each factor is unlikely to be consistent across all counties, we retained multiple components within each factor that explained about 70% of the variance for that factor. We also estimated prevalence based on the CDC Behavioral Risk Factor Surveillance System (BRFSS) COPD Model-based County Prevalence.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Face validity: Clinical Panel Review

During the clinical panel review both the Delphi and Nominal panel supported the measure with some concern. Other topics discussed by the panel included:

- Panelists advocated for restricting the indicator to patients 40 years of age and older and combining with asthma admissions in this age group. Empirical analysis confirmed that COPD diagnoses in cases under 40 years of age are rare. Panelists felt that combining these groups would eliminate the diagnostic uncertainly between asthma and COPD in older patients, and thus provide a cleaner measure.
- Smoking continuation/cessation may be a key component to disease progression in individuals with COPD. Panelists expressed mixed opinions on the ability of the healthcare system to affect smoking rates. Some noted that payer organizations may enhance coverage beyond current reimbursements available for smoking cessation efforts.

[5]Davies S, Romano PS, Schmidt EM, Schultz E, Geppert JJ, McDonald KM. Assessment of a Novel Hybrid Delphi and Nominal Groups Technique to Evaluate Quality Indicators. *Health services research.* 2011;46(6 Pt 1):2005-2018.

- As with all chronic conditions, comorbidities and disease severity are of concern. For COPD, other respiratory and cardiovascular conditions are of particular concern. Along with risk factors such as age, gender, race/ethnicity, socioeconomic status, and smoking rates, panelists emphasized that environmental factors may affect admissions rates for this indicator. These environmental factors include pollution levels, altitude, climate, and occupational exposures from local industries.
- Panelists also generally agreed that the high cost and complicated protocols for inhaler medications present major barriers to patient adherence to treatment recommendations. They further agreed that it is within the ability of the healthcare system to mitigate these barriers through efforts including offering high quality education on medication needs and inhaler use.
- Panelists felt this indicator may also reflect some amount of "social" hospital admissions. In other words, cases in which the physician determines social support or the home environment are insufficient for recovery outside of the hospital.
- The presence of observation units may impact admission rates for COPD.

Empirical Validity

In a negative binomial model (see section 2b2.2), health behaviors (HB) was the only statistically significant predictor (p<.00010), while SES/Environment was significant at a relaxed 0.10 threshold (p=0.09) (see Table 3a). Access to care (AC) and prevalence were not significant when HB and SES/E are included in the model. However, categorizing the variables into interpretable groups (HB, AC, and SES/E) resulted in significant collinearity in the models. In particular, there was a correlation of magnitude 0.77 between HB and one of the components for SES/E and correlations of magnitude between 0.40 and 0.50 between the components for AC and the other two components for SES/E (see Table 3b). Hence the relative importance of those factors should be interpreted with caution.

Table 3a: P-values from negative binomial model for PQI 05 as a function of prevalence and the various groups of principal components

	Prevalence	Health Behaviors	Access to Care	SES/Environment
PQI 05	0.14	<.0001	0.065	0.0032

Table 3b: Correlation matrix of principal components (PC) used in negative binomial models.

	HB	AC_1	AC_2	SES_1	SES_2	SES_3
Health Behaviors(HB)	1.00	-0.20	-0.13	0.22	**-0.77**	0.18
Access to Care (AC_1)	-0.20	1.00	0.00	**0.50**	0.13	**0.40**
Access to Care (AC_2)	-0.13	0.00	1.00	**0.44**	0.24	0.24
Socio-Economic Status (SES_1)	0.22	-0.50	0.44	1.00	0.00	0.00
Socio-Economic						

Status (SES_2)	-0.77	0.13	0.24	0.00	1.00	0.00
Socio-Economic Status (SES_3)	0.18	0.40	0.24	0.00	0.00	1.00

Note that the groups of variables corresponding to health behaviors (HB) were reduced to one principal component, access to care (AC) to two, and SES to three. Correlations with magnitude larger than 0.4 are bolded.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

While the individual health behaviors factor was the most predictive factor, moderate correlation between socioeconomic status and individual risk factors and access to care suggest that it is difficult to fully ascertain the unique contribution of each factor on COPD hospitalizations. The disparities table (see Table 2 in the supplemental files) demonstrates that zip codes in the highest income quartile have 57% lower admission rates than those in the lowest income quartile.

From a population health perspective such disparities argue for the importance of the indicator in capturing poor outcomes for vulnerable populations. Further, taking a population health perspective, efforts to decrease individual risk factors such as obesity, smoking and limited physical exercise (the variables included in the health behaviors factor) may decrease prevalence and exacerbations of COPD.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Using the 2013 data from 40 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

A total of 682 discharges were excluded due to diagnoses of cystic fibrosis and anomalies of the respiratory system. Removing this exclusion would increase the numerator count by 0.10%. The denominator does not change. Although discharges transferred into a hospital are excluded, these encounters are captured in the area level numerator via the originating hospitalization.

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e.*, *the value outweighs the burden of increased data collection and analysis.* *Note*: ***If patient preference is an exclusion***, *the measure must be specified so that the*

effect on the performance score is transparent, e.g., scores with and without exclusion)

The exclusion of cystic fibrosis and anomalies of the respiratory system has been retained to increase the face validity of the measure and to align the measure with the young adult and pediatric PQI/PDI for asthma admissions, although the impact on the numerator is minor for PQI 05. Patients meeting the exclusion are identifiable without additional burden using the same data as is used to identify numerator qualifying discharges.

The indicator excludes patients with severe chronic respiratory diseases, because COPD/asthma in the context of these diseases differs clinically from the patients with COPD/asthma alone.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with** 3 **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Sociodemographic Factors

The risk model incudes age and gender of the population as COPD prevalence and severity increase with age.

We considered multiple income risk factors for inclusion in the model, including % of households under the federal poverty level, household income, median % of poverty level (e.g. 200% of federal poverty level). The Public Health Disparities Geocoding Project has completed extensive evaluation of alternative income variables and has demonstrated that the percent poverty variable consistently detects expected gradients in heath across health outcomes, is widely available, and has a low rate of missing data even at the census tract level[6]. Our team has explored alternative income variables that do not outperform the poverty level variable (data not shown).

We also considered a conceptual model that acknowledged the impact of community factors such as clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise along with community norms and beliefs. However, these are factors that are difficult to measure within the framework of the AHRQ QIs, i.e., use with administrative data. These community factors can impact prevention of COPD and self-care for COPD, which in turn can impact hospitalization rates. Poverty can be a mitigating factor, inasmuch as impoverished communities are more likely to

[6] Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian S. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures-the public health disparities geocoding project. *American journal of public health.* 2003;93(10):1655-1671.

experience housing and food insecurity[7], air pollution[8], occupational exposure and have higher smoking rates. The relationship between environment, income and health is complex and the mechanism is not fully understood.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The process to select risk factors is described in the AHRQ QI Empirical Methods report. The results of the analyses are provided in the PQI Parameter Estimates document. Both documents are available to reviewers in the supporting materials. The results of the analyses are provided in the tables below as well as on the submitted excel spreadsheet.

There are several steps involved in estimating the QI risk-adjustment models.

1. Construct candidate covariates
2. Select model covariates
3. Estimate the models
4. Evaluate the models

Covariates are coded for each discharge record based on the data elements, data values, and logic described in the technical specifications and the appendices of the risk-adjustment coefficient tables. For a given covariate, if the discharge meets the technical specification for that covariate a value of "1" is assigned to the discharge level covariate data element. Otherwise a value of "0" is assigned to the discharge level covariate data element.

Table 4a. Risk Adjustment Coefficients for PQI05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
Intercept		1	-4.4435671	0.0065304	462995.27	<.0001
SEX	Female	1	-0.1787590	0.0082845	465.58353	<.0001
AGE	Male, Age 40-44	1	-2.9086319	0.01389835	43797.643	<.0001
AGE	Male, Age 45-49	1	-2.3056261	0.01113389	42882.855	<.0001
AGE	Male, Age 50-54	1	-1.7601027	0.00935114	35427.964	<.0001
AGE	Male, Age 55-59	1	-1.3778293	0.00869732	25096.882	<.0001
AGE	Male, Age 60-64	1	-1.1171939	0.00853497	17133.741	<.0001
AGE	Male, Age 65-69	1	-0.7704599	0.00833865	8537.0644	<.0001
AGE	Male, Age 70-74	1	-0.4071984	0.00832024	2395.1878	<.0001
AGE	Male, Age 75-79	1	-0.1945187	0.00858411	513.48951	<.0001
AGE	Male, Age 80-84	1	-0.0615533	0.00903698	46.393390	<.0001
AGE	Male, Age 85+		Referent	.	.	.
AGE	Female, Age 40-44	1	0.9300696	0.01700878	2990.0897	<.0001
AGE	Female, Age 45-49	1	0.8553427	0.01380390	3839.5153	<.0001
AGE	Female, Age 50-54	1	0.7027058	0.01179230	3550.9950	<.0001
AGE	Female, Age 55-59	1	0.5389804	0.01110700	2354.7902	<.0001
AGE	Female, Age 60-64	1	0.4310054	0.01097316	1542.7714	<.0001
AGE	Female, Age 65-69	1	0.3868430	0.01073165	1299.3807	<.0001
AGE	Female, Age 70-74	1	0.3277437	0.01071384	935.79004	<.0001
AGE	Female, Age 75-79	1	0.2511653	0.01106910	514.86640	<.0001
AGE	Female, Age 80-84	1	0.1565196	0.01163827	180.86752	<.0001
AGE	Female, Age 85+		Referent	.	.	.
c-statistic=0.5333						

7 Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the US. American journal of preventive medicine. 2009;36(1):74-81. e10.

8 Bell ML, Ebisu K. Environmental inequality in exposures to airborne particulate matter components in the United States. *Environmental health perspectives.* 2012;120(2):1699-1704.

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 4b. SES Risk Adjustment Coefficients for PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCEPT		1	-4.689609709	0.007776001	363714.6634	<.0001
SEX	Female	1	-0.179450471	0.008286004	469.0281238	<.0001
AGE	Male, Age 40-44	1	-2.910815144	0.013899371	43857.00701	<.0001
AGE	Male, Age 45-49	1	-2.304965504	0.01113517	42848.45217	<.0001
AGE	Male, Age 50-54	1	-1.758605166	0.009352584	35356.83854	<.0001
AGE	Male, Age 55-59	1	-1.378557835	0.008698813	25114.81296	<.0001
AGE	Male, Age 60-64	1	-1.120781087	0.008536476	17237.89952	<.0001
AGE	Male, Age 65-69	1	-0.77498984	0.008340149	8634.643968	<.0001
AGE	Male, Age 70-74	1	-0.41378771	0.008321873	2472.367231	<.0001
AGE	Male, Age 75-79	1	-0.202813362	0.008585853	557.9897283	<.0001
AGE	Male, Age 80-84	1	-0.065908284	0.009038621	53.17110854	<.0001
AGE	Male, Age 85+		Referent	.	.	.
AGE	Female, Age 40-44	1	0.931373981	0.017009612	2998.191762	<.0001
AGE	Female, Age 45-49	1	0.856478977	0.013804953	3849.138819	<.0001
AGE	Female, Age 50-54	1	0.702163477	0.011793526	3544.778653	<.0001
AGE	Female, Age 55-59	1	0.537224428	0.011108346	2338.90572	<.0001
AGE	Female, Age 60-64	1	0.429486868	0.010974569	1531.527722	<.0001
AGE	Female, Age 65-69	1	0.385477287	0.010733148	1289.862266	<.0001
AGE	Female, Age 70-74	1	0.326009411	0.010715482	925.6287226	<.0001
AGE	Female, Age 75-79	1	0.249326608	0.011070875	507.193115	<.0001
AGE	Female, Age 80-84	1	0.154204543	0.011640204	175.4983941	<.0001
AGE	Female, Age 85+		Referent	.	.	.
POVCAT	Poverty Decile 2	1	-0.034911688	0.006269083	31.01228056	<.0001
POVCAT	Poverty Decile 3	1	0.113567346	0.006034491	354.1812896	<.0001
POVCAT	Poverty Decile 4	1	0.171926509	0.005904751	847.7787023	<.0001
POVCAT	Poverty Decile 5	1	0.289646329	0.005759675	2528.949461	<.0001
POVCAT	Poverty Decile 6	1	0.281914805	0.005798991	2363.365447	<.0001
POVCAT	Poverty Decile 7	1	0.276491239	0.005852959	2231.576783	<.0001
POVCAT	Poverty Decile 8	1	0.342240605	0.005829532	3446.637713	<.0001
POVCAT	Poverty Decile 9	1	0.377000891	0.005832029	4178.740734	<.0001
POVCAT	Poverty Decile 10 (Highest percent poverty)[1]	1	0.594236298	0.00558918	11303.73997	<.0001
c-statistic=0.5328						

[1]Deciles are based on the percentage of households under the federal poverty level (FPL).

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

See above results from validity testing in Section 2b2.3.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest (i.e. admission for COPD). The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic. The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model based on the value of the observations covariates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. This creates a set of pairs of observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the non-event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common "goodness of fit" tests because these tests tend to not be informative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Table 5a. Age-sex Risk adjustment Model Discrimination and Calibration, for PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	24,552,754	0.000974	0.000975
2	22,924,527	0.001882	0.001912
3	21,816,496	0.002901	0.002841
4	19,456,360	0.003809	0.003915
5	17,621,075	0.004757	0.004842
6	13,729,288	0.006349	0.006055
7	9,661,011	0.00858	0.008844
8	7,269,004	0.009699	0.009779
9	7,412,410	0.010483	0.010243
10	4,358,167	0.011253	0.011189
C-Statistic	0.5333		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 5b. Age-sex and SES Risk adjustment Model Discrimination and Calibration, for PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	26,280,426	0.000954	0.000927
2	23,877,056	0.001924	0.001869
3	21,518,472	0.002916	0.002802
4	19,812,043	0.003869	0.003925
5	16,696,168	0.00497	0.005146
6	12,504,946	0.006539	0.006732
7	8,748,627	0.008095	0.008754
8	8,456,897	0.009571	0.009606
9	6,936,524	0.010927	0.010604
10	3,969,933	0.013314	0.011851
C-Statistic	0.5328		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

See Table 5 in 2b4.6

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

See Table 5 in 2b4.6

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated, as the observed to predicted values across the deciles range between 0.95– 1.03. The discrimination is low with a c-statistic of 0.53, presumably due to the limited predictors included. Addition of SES results in similar performance with observed to predicted values ranging across the deciles range between 0.89 – 1.08. The c-statistic does not change.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

This analysis assesses the probability that a county is higher or lower than a benchmark or threshold, given county size.

It reflects whether the indicator can discriminate the best performing counties from the lower performing counties.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of counties have a lower admission rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance).

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across counties of various sizes. Each county is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each county the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Counties are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each county to determine if the county rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 6 reports the proportion of counties above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of counties not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 6a. Performance Categories by County Size Decile PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Population Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	314	1,553.30	0.0000	0.7038	0.2962	0.7038	0.0255	0.2707
2	314	3,725.40	0.0000	0.7325	0.2675	0.7134	0.0955	0.1911
3	314	5,819.10	0.0000	0.7452	0.2548	0.6911	0.1146	0.1943
4	314	8,390.40	0.0000	0.8344	0.1656	0.6561	0.1911	0.1529
5	314	11,299.90	0.0000	0.8025	0.1975	0.6529	0.1879	0.1592
6	314	15,572.90	0.0096	0.8439	0.1465	0.6561	0.2134	0.1306
7	314	21,779.20	0.0955	0.8471	0.0573	0.6497	0.2197	0.1306

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Population Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
8	314	34,147.80	0.1051	0.8503	0.0446	0.7006	0.1752	0.1242
9	314	65,115.80	0.1306	0.8439	0.0255	0.7675	0.1433	0.0892
10	314	306,522.70	0.0955	0.8790	0.0255	0.8758	0.0860	0.0382
Overall	3,140	47,392.70	0.0436	0.8083	0.1481	0.7067	0.1452	0.1481

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 6b. SES Performance Categories by County Size Decile PQI 05 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Population Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	314	1,553.30	0.0000	0.5510	0.4490	0.6465	0.0318	0.3217
2	314	3,725.40	0.0000	0.6592	0.3408	0.6815	0.1019	0.2166
3	314	5,819.10	0.0000	0.6943	0.3057	0.6624	0.1369	0.2006
4	314	8,390.40	0.0032	0.7962	0.2006	0.5987	0.2006	0.2006
5	314	11,299.90	0.0701	0.7548	0.1752	0.6338	0.1975	0.1688
6	314	15,572.90	0.0955	0.8153	0.0892	0.6369	0.2229	0.1401
7	314	21,779.20	0.1146	0.8185	0.0669	0.6178	0.2548	0.1274
8	314	34,147.80	0.1210	0.8248	0.0541	0.6815	0.2038	0.1146
9	314	65,115.80	0.1401	0.7930	0.0669	0.7484	0.1752	0.0764
10	314	306,522.70	0.1115	0.8248	0.0637	0.8503	0.1146	0.0350
Overall	3,140	47,392.70	0.0656	0.7532	0.1812	0.6758	0.1640	0.1602

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

This indicator has strong discrimination to identify low performing counties for most counties; 85% of counties can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold). The indicator has strong discrimination, particularly for moderate to large counties to identify high performing counties; 85% of counties can be classified as better or worse than the benchmark. Performance discrimination remains strong when adding SES to risk adjustment.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PQI 05 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all county discharge records. The rate of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

For these variables, rates of missing data are typically less than 1% of the state database. It is unlikely the bias would occur from such a low rate of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases for missing data is appropriate.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements

and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because the indicator is based on readily available administrative billing and claims data and U.S. Census data, feasibility is not an issue.
The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees. Software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2011/Methodology.html CMS Medicaid Adult Core Measures http://www.medicaid.gov/medicaid-chip-program-information/by-topics/quality-of-care/adult-health-care-quality-measures.html Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings Maine Health Data Organization (MHDO), MONAHRQ Website http://gateway.maine.gov/mhdo/monahrq/Methodology.html Nevada Compare Care, MONAHRQ website http://nevadacomparecare.net/ Oklahoma State Department of Health, MONAHRQ https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html Utah Department of Health, MONAHRQ website https://health.utah.gov/myhealthcare/monahrq/ Virginia Health Information, MONAHRQ website http://www.vhi.org/MONAHRQ/default.asp?yr=2013 Washington State, MONAHRQ website http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Defi

	nitions California Office of Statewide Health Planning and Development, Healthcare Information Division http://oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/ Connecticut, Office of Health Care Access http://www.ct.gov/dph/lib/dph/ohca/publications/2010/prev_hosp_report01-2010.pdf St John´s Episcopal Hospital http://www.ehs.org/documents/ST.-JOHNS-EPISCOPAL-HOSPITAL-COMMUNITY-SERVICE-PLAN.pdf Arkansas Department of Human Services: Arkansas Medicaid Performance http://humanservices.arkansas.gov/dms/Pages/aqg-Report-Methodology.aspx#Quality Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW) http://www.healthindicators.gov/Resources/Initiatives/CMS/Prevention-Quality-Indicators-Report_20/Indicator/Report Northwest Hospital and Medical Center http://www.nwhospital.org/downloads/pdfs/Northwest-Hospital-CHNA-2013.pdf Payment Program CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use Reports (QRUR) http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2011-ACSC-Outcomes-Measures.pdf CMS Medicare Shared Savings Program https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html Oregon Health Authority http://www.oregon.gov/oha/analytics/Pages/CCO-Baseline-Data.aspx Regulatory and Accreditation Programs Statewide Quality Advisory Committee (Massachusetts) http://chiamass.gov/sqms/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) West Jefferson Medical Center http://www.wjmc.org/docs/WJMC-Secondary-Data-Profile-09-23-2013.pdf

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting:
Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona.
http://pub.azdhs.gov/hospital-discharge-stats/2011/Methodology.html

CMS Medicaid Adult Core Measures1
http://www.medicaid.gov/medicaid-chip-program-information/by-topics/quality-of-care/adult-health-care-quality-measures.html
ACA mandated reporting system, in which Medicaid states can voluntarily report on a set of core measures.
1The numerator of PQI 05 is used in this program. The denominator and risk adjustment is modified to meet specific program requirements.

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website

Hospital quality ratings from all hospitals in Connecticut.
http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine.
http://gateway.maine.gov/mhdo/monahrq/Methodology.html

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada: Quality reporting on hospitals across the state of Nevada Under NV Regulation R151-8 this transparency website presents hospital quality and utilization information.
http://nevadacomparecare.net/

Oklahoma State Department of Health, MONAHRQ
Compares quality ratings on hospitals across Oklahoma.
https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html

Utah Department of Health, MONAHRQ website
Hospital quality ratings from all hospitals in Utah.
https://health.utah.gov/myhealthcare/monahrq/

Virginia Health Information, MONAHRQ website
Compares quality ratings on hospitals across Virginia.
http://www.vhi.org/MONAHRQ/default.asp?yr=2013

Washington State, MONAHRQ website
Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals.
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions

California Office of Statewide Health Planning and Development, Healthcare Information Division
OSHPD Patient Discharge Data from all hospitals in California, totaling over 4 million records annually.
http://oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/

Connecticut, Office of Health Care Access
Preventable Hospitalizations in Connecticut: A Current Assessment of Access to Community Health Services: 2004-2009 state- and county-level hospital admission rate data from most hospitals in CT.
http://www.ct.gov/dph/lib/dph/ohca/publications/2010/prev_hosp_report01-2010.pdf

St John´s Episcopal Hospital
Community Health Needs Assessment (CHNA) required by federal law and Patient Protection and Affordable Health Care Act.
http://www.ehs.org/documents/ST.-JOHNS-EPISCOPAL-HOSPITAL-COMMUNITY-SERVICE-PLAN.pdf

Arkansas Department of Human Services: Arkansas Medicaid Performance
Arkansas state Department of Human Services with use of Medicaid funds for children and elderly.
http://humanservices.arkansas.gov/dms/Pages/aqg-Report-Methodology.aspx#Quality

Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
Purpose of the HIW is to: Provide a single, user-friendly, source for national, state, and community health indicators; Facilitate harmonization of indicators across initiatives; Link indicators with evidence-based interventions.
http://www.healthindicators.gov/Resources/Initiatives/CMS/Prevention-Quality-Indicators-Report_20/Indicator/Report

Northwest Hospital and Medical Center
Community Health Needs Assessment (CHNA) required by federal law and Patient Protection and Affordable Health Care Act.
http://www.nwhospital.org/downloads/pdfs/Northwest-Hospital-CHNA-2013.pdf

Quality Improvement:
West Jefferson Medical Center
Reports indicators of potentially avoidable hospitalizations associated with the parish in which it is located, and compared those

indicators with state-level indicators.
http://www.wjmc.org/docs/WJMC-Secondary-Data-Profile-09-23-2013.pdf

Regulatory and Accreditation Programs:
Statewide Quality Advisory Committee (Massachusetts)
The committee annually recommends a standard set of health metrics to use throughout statewide health quality efforts.
http://chiamass.gov/sqms/

Payment Programs:
CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use Reports (QRUR) Program includes measures of Ambulatory Care Sensitive Conditions (ACSC), used by Physicians receiving Medicare FFS payment modifiers.
http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2011-ACSC-Outcomes-Measures.pdf

Oregon Health Authority
Coordinated Care Organization (CCO) implementing Oregon´s pay for performance program using quality health metrics.
http://www.oregon.gov/oha/analytics/Pages/CCO-Baseline-Data.aspx

CMS Medicare Shared Savings Program1
https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html
ACO program uses quality measures to establish a performance standard to qualify for receipt of a share of savings.
1The numerator of PQI 05 is used in this program. The denominator and risk adjustment is modified to meet specific program requirements.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
n/a

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
n/a

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

See Table 1 in response to question 1b.2. The rate of PQI 05 hospital admissions has decreased from 2011 to 2013. In 2012 the rate was 5.2 per 1,000 while in 2013 the rate was 4.4 per 1,000. This decrease represents over 104,000 fewer hospitalizations. Further, the variation between counties decreased substantially in 2013, although we cannot determine whether this is a single-year anomaly or an actual trend. Additionally, it is important to note, the standard deviation is highly sensitive to outliers and the observed change appeared to be due to outlier counties.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

During a structured clinical panel review, panelists postulated that some uses of this indicator could disincentive care for high risk individuals. However, no evidence of this unintended consequence has arisen during actual use of the indicator. Rather, identification of high rates can help to target populations most in need of intervention.

Panelists in the same structured review and subsequent expert panel review noted that treatment of COPD in observation care may substitute for inpatient treatment, that this substitution may be systematic between areas and that this will impact the rate of the indicator. During a literature review, we identified no studies that specifically examined observation stays as a substitute for inpatient care. In a retrospective analysis of a 2002-2011 large administrative claims database of commercially insured individuals in the USA, COPD did not appear in the most frequent diagnosis categories in either emergency department-based or inpatient-based observation units.1 A retrospective analysis of observation stays from three distinct data sources: 2010 Atlanta hospitals protocol driven observation units, 2010 Georgia hospitals for observation units (including protocol-driven, discretionary care and all bed locations), and 2009-10 National Hospital Ambulatory Medical Care Survey (NHAMCS) for similarly diverse of observation units found that COPD/asthma ranked the most common condition managed in observation services. However the study did not examine diagnoses by age group.2

1. Overman RA, Freburger JK, Assimon MM, Li X, Brookhart MA. Observation stays in administrative claims databases: underestimation of hospitalized cases. Pharmacoepidemiology and drug safety. Sep 2014;23(9):902-910.
2. Ross MA, Hockenberry JM, Mutter R, Barrett M, Wheatley M, Pitts SR. Protocol-driven emergency department observation units offer savings, shorter stays, and reduced admissions. Health Aff (Millwood). Dec 2013;32(12):2149-2156.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PQI05_NQF0275_Supplemental_Files_151214.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Carol, Stocks, Carol.Stocks@ahrq.hhs.gov
Co.3 **Measure Developer if different from Measure Steward:** Agency for Healthcare Research and Quality
Co.4 **Point of Contact:** Carol, Stocks, Carol.Stocks@ahrq.hhs.gov

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The following panelists participated in a 2009 structured panel review of the Agency for Healthcare Research and Quality Prevention Quality Indicators, which focused on evaluating expansion of the indicators to alternative denominator populations. The panel used a modified Delphi approach to evaluate the indicators, using a method that combined a nominal group technique and a Delphi technique.1 All panelists rated the indicators and received feedback from other panelists. The panelists participated in a conference call to discuss the indicators and the discussion was summarized and distributed to the group before final rating. Some panelists requested that their affiliation with this report remain anonymous, and this list is therefore a partial representation of the individuals that comprised the panels in their entirety.

1. Davies S, McDonald KM, Schmidt E, Geppert J, Romano PS. Expanding the uses of AHRQ's Prevention Quality Indicators: Validity from the clinician perspective. Med Care. Aug 2011; 49(8): 679-685.

Sandra G. Adams, MD, MS, FCCP
Pulmonolary & Critical Care Medicine
South Texas Veterans Health Care System
University of Texas Health Science Center
San Antonio, Texas
Nominated by American College of Chest Physicians

Wilbert S. Aronow, MD, FACC, FAHA, AGSF, FCCP
Geriatric Medicine, Cardiology
Westchester Medical Center
New York Medical College
Valhalla, New York
Nominated by The American Geriatrics Society

John D. Birkmeyer, MD, FACS

Surgery
University of Michigan Hospital
Ann Arbor, Michigan
Nominated by American College of Surgeons

James H. Black, III, MD
Vascular Surgery
Johns Hopkins Hospital
Johns Hopkings University
Baltimore, Maryland
Nominated by Society for Vascular Surgery

Cynthia Boyd, MD
Geriatric Medicine
Johns Hopkins Hospital, Bayview Medical Center
Johns Hopkins University
Batlimore, Maryland
Nominated by The American Geriatrics Society

Suzanne Bradley, MD
Geriatric Medicine and Infectious Diseases
Veterns Affairs Ann Arbor Healthcare System
University of Michigan
Ann Arbor, Michigan
Nominated by The American Geriatrics Society

Vickie Brown, RN, MPH, CIC
Hospital Epidemiology
University of North Carolina Hospital
Chapel Hill, North Carolina
Nominated by Association for Professionals in Infection Control and Epidemiology

John Buckley, MD, MPH
Pulmonary & Critical Care Medicine
Henry Ford Hospital
Wayne State University
Detroit, Michigan
Nominated by Society of General Internal Medicine

Pauline M. Camacho, MD, FACE
Division of Endocrinology and Metabolism
Loyola University Medical Center
Loyola Stritch School of Medicine
Maywood/Chicago, Illinois
Nominated by American Academy of Clinical Endocrinologists

Dickson Cheung, MD, MBA, MPH
Emergency Medicine
Sky Ridge Medical Center
Lone Tree, Colorado
Nominated by American College of Emergency Physicians

Leslie Davis, MSN, RN, ANP-C
Division of Cardiology
Department of Medicine
University of North Carolina
Chapel Hill, North Carolina

Nominated by American Academy of Nurse Practitioners

Barbara DeBaun, MSN, RN, CIC
Improvement Advisor
Beacon, Bay Area Patient Safety Collaborative
Bay Area Counties, California
Nominated by Association for Professionals in Infection Control and Epidemiology

Gregory J. Dehmer, MD, FACC, FSCAI
Interventional Cardiology
Scott & White Healthcare
Texas A&M University Health Science Center
Temple, Texas
Nominated by American College of Cardiology

Shawkat Dhanani, MD, MPH
Geriatric Medicine
Veterans Affairs Greater Los Angeles Healthcare System
University of California at Los Angeles
Los Angeles, California
Nominated by The American Geriatrics Society

Michelle Farber, RN, CIC
Infection Prevention and Control
Mercy Community Hospital
Coon Rapids, Minnesota
Nominated by Association for Professionals in Infection Control and Epidemiology

Amy Fendrich, MD
Internal Medicine
Memorial Regional Hospital Primary Care Clinics
Dania Beach, Florida
Nominated by American Puclic Health Association, Medical Care Section

Carlos M. Ferrario, MD, FAHA, FASA, FACC
Cardiology
Wake Forest University Health Science Center
Winston-Salem, North Carolina
Nominated by American College of Cardiology

John E. Gardella, MD, MBA, FCCP, FHM
Hospital Medicine
Presbyterian Hospital, Novant Health
Charlotte, North Carolina
Nominated by Society of Hospital Medicine

Eric Gertner, MD, MPH, FACP
Internal Medicine
Lehigh Valley Hospital
Penn State College of Medicine
Allentown, Pennsylvania
Nominated by American College of Physicians

James M. Gill, MD, MPH
Family Medicine
Delaware Valley Outcomes Research
Jefferson Medical College

Wilmington, Delaware
Nominated by American Puclic Health Association, Medical Care Section

Louis Gilleran, MD, MPH, FACPM
Preventive Aerospace Medicine
Naval Medical Center San Diego
San Diego, California
Nominated by American Puclic Health Association, Medical Care Section

Christopher Gonzalez, MD, MBA
Urology
Northwestern University Medical Center
Chicago, Ilinois
Nominated by American Urological Association

Michael K. Gould, MD, MS
Pulmonary and Critical Care Medicine
Veterans Affairs Palo Alto Health Care System
Stanford University
Palo Alto, California
Nominated by American Thoracic Society

George Grunberger, MD, FACP, FACE
Diabetes, Endocrinology
Grunberger Diabetes Institute
Wayne State University
Bloomfield Hills & Detroit, Michigan
Nominated by American Academy of Clinical Endocrinologists

A. Seiji Hayashi, MD, MPH
Family Medicine
Unity Health Care, Inc.
George Washington University
Washinton, DC
Nominated by American Puclic Health Association, Medical Care Section

Richard Hellman, MD, FACP, FACE
Clinical Endocrinology
North Kansas City Hospital
University of Missouri, Kansas City School of Medicine
Kansas City, Missouri
Nominated by American Academy of Clinical Endocrinologists

Steve J. Hodges, MD
Surgery, Pediatric Urology
Wake Forest University Health Sciences
Institute for Regenerative Medicine
Winston-Salem, North Carolina
Nominated by American College of Surgeons (by proxy through Dr. Anthony Atala)

Mary Johnson, MS, RD, CDE, BC-ADM
Diabetes Quality and Education
Geisinger Health System
Central Pennsylvania
Nominated by American Dietetic Association

Jeanette Kalupa, MSN, ACNP-BC, APNP

Acute Care Nurse Practioner, Hospitalist
Cogent Healthcare
Aurora St. Luke´s Medical Center
Milwaukee, Wisconsin
Nominated by American Academy of Nurse Practitioners

Marjorie L. King, MD, FACC, FAACVPR
Cardiology
Helen Hayes Hospital
Columbia University
West Haverstraw, New York
Nominated by American College of Cardiology

Jerry A. Krishnan, MD, PhD
Pulmonary and Critical Care Medicine
University of Chicago, Asthma and COPD Center
Chicago, Ilinois
Nominated by American Thoracic Society

Geoffrey Lamb, MD
Internal Medicine
Froedtert Hospital
Medical College of Wisconsin
Milwaukee, Wisconsin
Nominated by Society of General Internal Medicine

Gene Lambert, MD, MBA
Hospital Medicine
Massachusetts General Hospital
Harvard University
Boston, Massachusetts
Nominated by Society of Hospital Medicine

David E. Lanfear, MD, MS, FACC
Cardiology
Henry Ford Hospital
Wayne State University
Detroit, Michigan
Nominated by American College of Cardiology

Karen Lasser, MD, MPH
Primary Care, Internal Medcine
Boston University, Medical Center
Boston, Massachusetts
Nominated by Society of General Internal Medicine

Susan Lee, MS, NP-C, FAANP
Family Medicine, Internal Medicine
Bloomington Hospital
Bloomington, Indiana
Nominated by American Academy of Nurse Practitioners

Edgar V. Lerma, MD
Nephrology
Associates in Nephrology, S.C.
University of Illinois at Chicago
Chicago, Illinois

Nominated by The American Society of Nephrology

James T. Li, MD, PhD
Internal Medicine, Allergy
Mayo Clinic, College of Medicine
Rochester, Minnesota
Nominated by American Academy of Allergy Asthma and Immunology

Jenny J. Lin, MD
Internal Medicine
Mount Sinai Medical Center
Mount Sinai Medical School
New York, New York
Nominated by Society of General Internal Medicine

Frank LoGerfo, MD
Vascular Surgery
Beth Israel Deaconess Hospital
Harvard Medical School
Boston, Massachusetts
Nominated by Society for Vascular Surgery

John J. Lopez, MD, FACC
Interventional Cardiology
Loyola University Medical Center
Stritch School of Medicine, Loyola University
Maywood, Illinois
Nominated by American College of Cardiology

Thomas D. MacKenzie, MD, MSPH
Internal Medicine
Denver Health
University of Colorado
Denver, Colorado
Nominated by Society of General Internal Medicine

William Marston, MD
Vascular Surgery
University of North Carolina
Chapel Hill, North Carolina
Nominated by Society for Vascular Surgery

Helmut Meisl, MD, FACEP
Emergency Medicine
Good Samaritan Hospital
San Jose, California
Nominated by American College of Emergency Physicians

Mark L. Metersky, MD, FCCP
Pulmonary and Critical Care Medicine
University of Connecticut, Health Center
Farmington, Connecticut
Nominated by American College of Chest Physicians

Lena M. Napolitano, MD, FACS, FCCM
Surgery
University of Michigan

Ann Arbor, Michigan
Nominated by American College of Surgeons

John A. Parker, Jr., MD
Family Medicine
Marshall University
Huntington, West Virginia
Nominated by American Academy of Family Physicians

Jay I. Peters, MD, FCCP
Chief Pulmonary & Critical Care Medicine
University of Texas Health Science Center
San Antonio, Texas
Nominated by American College of Chest Physicians

Michael P. Phelan, MD, FACEP
Emergency Medicine
Cleveland Clinic
Cleveland, Ohio
Nominated by American College of Emergency Physicians

Mark Potter, MD
Family Medicine
University of Illinois at Chicago, Medical Center
Chicago, Illinios
Nominated by American Academy of Family Physicians

Carol Rasmussen, MSN, NP-C, CDE
Diabetes Program
Exodus Healthcare
Magna, Utah
Nominated by American Association of Diabetes Educators

H. David Reines, MD, FACS
Surgery
Inova Fairfax Hospital
Virginia Commonwealth University
Falls Church, Virginia
Nominated by American College of Surgeons

Barry Saver, MD, MPH
Family Medicine
University of Massachusetts Memorial
Worcester, Massachusetts
Nominated by American Academy of Family Physicians

Douglas J.E. Schuerer, MD, FACS, FCCM
Acute and Critical Care Surgery
Barnes Jewish Hospital
Washington University
St. Louis, Missouri
Nominated by American College of Surgeons

June Schulz, RRT, FAACVPR
Respiratory Care, Pulmonary Rehabilitation
Sanford University of South Dakota Medical Center
University of South Dakota

Sioux Falls, South Dakota
Nominated by American Association of Cardiovascular and Pulmonary Rehabilitation

Kristine M. Thompson, MD
Emergency Medicine
Mayo Clinic
Jacksonville, Florida
Nominated by American College of Emergency Physicians

Francesca J. Torriani, MD, FIDSA
Infectious Disease, Epidemiology
University of California at San Diego
San Diego, California
Nominated by Infectious Disease Association of California

Dace L. Trence, MD, FACE
Division of Metabolism, Endocrinology, and Nutrition
University of Washington Medical Center
Seattle, Washington
Nominated by American Academy of Clinical Endocrinologists

Arjun K. Venkatesh, MD, MBA
Emergency Medicine
Brigham and Women´s Hospital, Massachusetts General Hospital
Harvard University
Boston, Massachusetts
Nominated by American College of Emergency Physicians

Raoul Wolf, MBBCh, FCCP, FAAAI
Allergy and Clinical Immunology
La Rabida Chidrens Hospital, Comer
University of Chicago
Chicago, Illinois
Nominated by American College of Chest Physicians

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2007

Ad.3 **Month and Year of most recent revision:** 11, 2007

Ad.4 **What is your frequency for review/update of this measure?** Annually

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.6 **Copyright statement:** The AHRQ QI software is publicly available. We have no copyright disclaimers.

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0279
Measure Title: Bacterial Pneumonia Admission Rate (PQI 11)
Measure Steward: Agency for Healthcare Research and Quality
Brief Description of Measure: Admissions with a principal diagnosis of bacterial pneumonia per 1,000 population, ages 18 years and older. Excludes sickle cell or hemoglobin-S admissions, other indications of immunocompromised state admissions, obstetric admissions, and transfers from other institutions.
Developer Rationale: This indicator is intended to identify hospitalizations for pneumonia, either specified as bacterial or unspecified organism. With access to high quality care, early intervention and appropriate pharmaceutical treatment this condition can often be managed on an outpatient basis.

Numerator Statement: Discharges, for patients ages 18 years and older, with a principal ICD-9-CM or ICD-10-CM-PCS diagnosis code for bacterial pneumonia.

[NOTE: By definition, discharges with a principal diagnosis of bacterial pneumonia are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QI software does not explicitly exclude obstetric cases.]
Denominator Statement: Population ages 18 years and older in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.
Denominator Exclusions: Not applicable.

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Population : County or City

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

- This is a population-level measure. The level of analysis is county.
- A systematic review of the body of evidence is not required for outcome measures
- The developer provides the following rationale for this outcome measure: Access to high quality care, early intervention and appropriate pharmaceutical treatment, may minimize the likelihood of milder respiratory conditions progressing to pneumonia, reducing the likelihood of hospitalizations.
- The developer identifies the following guidelines as providing processes to identify and treat community-acquired pneumonia in order to reduce hospitalization:
    - Infectious Disease Society of America (IDSA) http://www.idsociety.org/Organ_System/
    - American Thoracic Society (ATS) http://www.thoracic.org/statements/tuberculosis-pneumonia.php
- Although not required per NQF guidance, the developer conducted a literature review (January 2012-October 2015) related to aspects of hospitalization for pneumonia, as follows:
    - Costs and association with other outcomes
    - Risk factors
    - *Disparities*
    - *Access to care*
    - *Variation in Admission Practice Patterns*

Questions for the Committee:

- *Although the developer provides updated evidence related to aspects of hospitalization for pneumonia, does the Committee agree the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on Evidence?*
- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer reports:

- Bacterial pneumonia is a relatively common acute condition. In 2011, pneumonia was the second most common principal diagnosis recorded in U.S. hospital stays, accounting for approximately 3,818,000 stays. It is a leading cause of morbidity/mortality and causes high use of resources.

Reference Population Rate and Distribution of County Performance for PQI 11
Overall Reference Population Rate

Year	Number of Counties	Number of Events (Numerator)	Population at Risk (Denominator)	Observed Rate Per 1,000
2009	3,137	702,634	232,379,612	3.0236
2010	3,141	668,530	234,909,365	2.8459
2011	3,144	678,908	237,419,828	2.8595
2012	3,141	638,091	237,830,861	2.683
2013	3,140	550,294	240,482,275	2.2883

Distribution of County-level Observed Rates in Reference Population Per 1,000
(p=percentile)

Year	Number of Counties	Mean	SD	p5	p25	Median	p75	p95
2009	3,137	5.20	25.36	0.05	2.53	3.91	5.83	10.92

2010	3,141	5.26	30.93	0.05	2.38	3.78	5.57	10.06
2011	3,144	5.24	30.91	0.08	2.40	3.73	5.55	10.11
2012	3,141	4.59	30.89	0.10	2.03	3.27	4.74	8.25
2013	3,140	3.28	2.43	0.02	1.68	2.99	4.48	7.59

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Disparities

- The developer reports the following disparities: Male patients, patients 65 and over, patients with the lowest income, and patients living in rural locations have the highest rates.

Questions for the Committee:

o *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**It would appear that this measure does relate to desired outcomes by carefully measuring the relationship between access to high quality care, early intervention and appropriate pharmaceutical treatment with the degree to which the likelihood of milder respiratory conditions progressing to pneumonia, reducing the likelihood of hospitalizations are minimized.

- The developer identifies the following guidelines as providing processes to identify and treat community-acquired pneumonia in order to reduce hospitalization:

o Infectious Disease Society of America (IDSA) http://www.idsociety.org/Organ_System/

o American Thoracic Society (ATS) http://www.thoracic.org/statements/tuberculosis-pneumonia.php

- Although not required per NQF guidance, the developer conducted a literature review (January 2012-October 2015) related to aspects of hospitalization for pneumonia, as follows:

o Costs and association with other outcomes

o Risk factors

o Disparities

o Access to care

**The evidence provided indicated a direct outcomes measure. If better outpatient management of conditions occurred it would result in less hospitalizations for bacterial pneumonia.

**o Although the developer provides updated evidence related to aspects of hospitalization for pneumonia, does the Committee agree the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on Evidence? YES

Is there at least one thing that the provider can do to achieve a change in the measure results?

YES, With access to high quality care, early intervention and appropriate pharmaceutical treatment this condition can often be managed on an outpatient basis.

**The measure is for an outcome. It is related to several types of interventions including access to care, vaccination, and potentially differences in provider behavior.

**The developer provides some convincing data around variation in pneumonia admission rates related to quality (e.g., nh ownership status) thus do not feel there is need for repeat discussion.

**Yes health action identified and supports the outcome measure

** Low to moderate rating for evidence. The measure aims to evaluate admission rates, but has primary discharge diagnosis in numerator. Although a proxy for admission diagnosis, this will miss many cases and underestimate rate. Also, I'm not sure who this measure is for-providers, hospitals, others? As the developer states, this is a population health indicator with a wide range of factors, but who is being evaluated with this measure? The level of analysis is at the County or City level, but within this level, who is being measured. Finally, what evidence is there that knowledge of the admission rate leads to interventions that change outcomes or even that lead to treatment of the condition on an outpatient basis?

-One specific healthcare action, vaccination, is identified, but unrelated to measure rationale of admission rate (although there is

evidence that vaccination improves measure results).

1b. Performance Gap

Comments:

**Yes. It supports the developers thesis that bacterial pneumonia is a relatively common acute condition. In 2011, pneumonia was the second most common principal diagnosis recorded in U.S. hospital stays, accounting for approximately 3,818,000 stays. It is a leading cause of morbidity/mortality and causes high use of resources.

**The gap in care demonstrated a disparity in older men, lower socio-economic status and rural areas. This indicates that there is an opportunity to improve outcomes with a focus on these sub populations.

**Is there a gap in care that warrants a national performance measure?

The developer reports the following disparities: Male patients, patients 65 and over, patients with the lowest income, and patients living in rural locations have the highest rates.

**The data do support a performance gap, with differences between different groups. Disparities do exist, males, age over 65, lower income. Given the NMQF interest in adult vaccination rates one might ask if their is a disparity in this measure related to race.

**Yes. High

**data provided supports that there is room to avoid hospitalization of patients with bacterial pneumonia, particularly in the named disparate populations. These represent opportunities for providers to apply early intervention activities and minimize resulting admissions and lessen health care expenditures on said hospitaliations

** Performance gap noted as 2nd leading cause of hospital admissions. Observed rate seems high, but not sure what it is compared to other conditions. Performance seems to be improving over that last 5 years. Disparities noted among male, elderly, low income and rural populations.

1c. High Priority (previously referred to as High Impact)

Comments:

**n/a

**NA

**N/a

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Administrative claims

Specifications:

- The developer indicates some changes to the measure specifications since the last endorsement review:
  - Added ICD-9-CM diagnosis codes 452.40 Staphylococcal pneumonia NOS, 482.41 Methicillin susceptible staphylococcal pneumonia, 482.42 MRSA pneumonia, 482.49 Staphylococcal pneumonia NEC to numerator definition of pneumonia
  - Removed numerator exclusions: MDC14 (pregnancy, childbirth, and puerperium); MDC15 (newborn and other neonates). By definition, discharges with a principal diagnosis of bacterial pneumonia are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QITM software does not explicitly exclude obstetric cases.
  - Added exclusion of patients with any diagnosis code or procedure code for Immunocompromised state
  - The data upon which to base the reference population was updated with 2013 US Census population estimates
  - Fiscal Year coding updates

- The numerator for this measure is: *Number of discharges for patients ages 18 years and older, with a principal ICD-9-CM or ICD-10-CM-PCS diagnosis code for bacterial pneumonia*.
- The denominator for this measure is: *Population ages 18 years and older in metropolitan area or county*. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.
- This outcome measure is risk adjusted, using a statistical risk model.
- The calculation algorithm is stated in S.18.

Questions for the Committee :

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Summary of reliability testing is not available from the prior review.

Describe any updates to testing

- The developer indicates there are updates to the reliability testing since the last submission
  - Reliability testing at the level of the measure score has been conducted using more current data
  - Risk adjustment is included.
  - Data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) is used: 40 states representing about 89% of U.S. county hospital discharges, for about 30 million hospital discharges from community hospitals. Data were from 2009-2013
- The developer provides two risk models:
  - age and gender composition of the county; and
  - optional addition to these two variables of the percent of households falling below the federal poverty level

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Reliability testing was conducted at the performance measure level, but not the individual element level.
- Testing was conducted using signal-to-noise analysis assessing the reliability to confidently distinguish the performance among counties.
- Specifically, the signal-to-noise ratio refers to the entire population of U.S. counties, comparing the degree to which rates are different from county to county (the signal) to how stable the rates are within counties (the noise).

Results of reliability testing

- The developer reported a signal-to-noise ratio of 0.97, which the developer indicates is strong.
- The developer reported that when SDS is added to the risk adjustment the signal-to-noise ratio of 0.96.

Guidance from the Reliability Algorithm: 1 → 2 → 4 → 5 → 6 (highest eligible rating is HIGH)

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Summary of validity testing is not available from the prior review.

Describe any updates to validity testing:

- No new empirical validity testing at the measure score level was performed; as noted below, face validity was assessed.
- The developer provides updated information using more recent data for analyses related to the risk model, exclusions, and meaningful differences.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- The measure was tested for face validity with input from four clinical expert panels involving 73 panelists. The panel was convened from 2008-2009.

Validity testing results:

- The developer reports the panels indicated the measure was useful. Specific actions could improve rates, such as access to medications, reduction of risk factors, such as COPD/asthma, diabetes and smoking and increasing vaccination in vulnerable populations. The developer and panels acknowledged complex factors influence the measure.

Question for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer excludes cases with missing gender, age, quarter, year, principal diagnosis or county. The developer states these exclusions never exceed 1% of eligible records.
- 1,783 discharges were excluded due to diagnoses of sickle cell disease. Removing this exclusion would increase the numerator count by 0.27%. The exclusion for sickle cell has been retained to increase the face validity of the measure and to align the measure with the pediatric PDI for pneumonia admissions. Patients are identified without additional burden.

- 69,135 discharges were excluded due to diagnoses of immunocompromised state. Removing this exclusion would increase the numerator count by 10.5%. Pneumonias in special populations with immunocompromised states are excluded since infections in these populations may be more likely to progress or require hospitalization despite access to high quality care.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

- The developer provides two risk models:
  - age and gender composition of the county; and
  - optional addition to these two variables of the percent of households falling below the federal poverty level

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

The developer reports:

- The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect)
  - The covariates are gender and age (in 5-year age groups)
  - An option model is available that includes percent of households under the federal poverty level
  - A conceptual model acknowledging the impact of community factors also was considered (e.g., clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise, community norms and beliefs, etc., which can impact hospitalization rates).
- The risk adjustment model was calibrated by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.
  - The developer reports the observed to predicated values across the deciles range between 0.99-1.01, indicating it is well calibrated.
  - The developer reports the c-statistic (measure of well the risk adjustment model distinguishes events from non-events, is the c-statistic) was 0.58, which it concludes was poor and presumes was due to the limited predictors included. The developer noted the addition of SES to the model did change performance.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Were appropriate community factors considered in the conceptual model? Should race be considered in the risk model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer assessed the probability that a county is higher or lower than a benchmark or threshold, given county size—i.e., whether the indicator can discriminate the best performing counties from the lower performing counties. The developer reported:

- The developer reports on discrimination results for counties.
- The developer indicates the "event" is hospitalization for bacterial pneumonia.

The developer reports:

- Identify low performing counties for most counties: 78% of counties can be classified as better or worse than

the threshold

- Identify high performing counties for moderate to large counties; 10% of counties can be classified as better or worse than the benchmark.

Question for the Committee:

o *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

- The AHRQ QIs use frequently reported administrative data variables. PQI 11 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis.
- Missing data handled – rates are typically less than 1% of the state database.

Guidance from the Validity Algorithm: 1 → 2 → 3 → 4 → 5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The specifications do not appear to be inconsistant with the evidence. In fact, the developer reports specifications to be consistent with evidence presented in 1a.

**The specifications are consistent with the evidence and the target population.

**Are the specifications consistent with the evidence? YES

**Face validity only has been measured. While it is likely that the measure is reflective of "quality", it is not straightforward.

Note - The developer reports the panels indicated the measure was useful. Specific actions could improve rates, such as access to medications, reduction of risk factors, such as COPD/asthma, diabetes and smoking and increasing vaccination in vulnerable populations. The developer and panels acknowledged complex factors influence the measure.

**Yes

**No concerns

** Measure specifications are clearly defined. However, the measure aims to evaluate admission rates, but has primary discharge diagnosis in numerator. Although a proxy for admission diagnosis, this will miss many cases and underestimate rate. Risk adjustment is calculated through a statistical risk model. Measure likely to be implemented consistently.

2a2. Reliability Testing

Comments:

**The measure was tested for face validity with input from four clinical expert panels involving 73 panelists. The panel was convened from 2008-2009.

The developer reports the panels indicated the measure was useful. Specific actions could improve rates, such as access to medications, reduction of risk factors, such as COPD/asthma, diabetes and smoking and increasing vaccination in vulnerable populations. The developer and panels acknowledged complex factors influence the measure.

**The validity testing indicated a sufficient validity so that conclusions could be made. It demonstrated that specific actions could improve outcomes.

**Do the results demonstrate sufficient validity so that conclusions about quality can be made? the clinical panel reviewing the use of this PQI supported the measure with some concern (related to confounders such as access to vaccines, questions regarding its reflection of access rather than quality of care, patient factors limiting control of system over admission rates, role of co-morbidities, use/over-use of Abx, social admissions). panel identified specific actions that could improve rates, especially from a population health perspective, such as access to medications, reduction of risk factors, such as COPD/asthma, diabetes and smoking and increasing vaccination in vulnerable populations.

**The developer reports the panels indicated the measure was useful. Specific actions could improve rates, such as access to medications, reduction of risk factors, such as COPD/asthma, diabetes and smoking and increasing vaccination in vulnerable populations. The developer and panels acknowledged complex factors influence the measure.

**Yes, rating: Low

**With the availability of performance data readily available, would expect that validity of the measure could be easily tested using computed performance scores rather than just the historical face validity testing. Additionally, there were no additional validity testing efforts to validate the new exclusion (immunocompromised), or change in age (removing the 18-39 population)

** Summary of reliability testing is not available from prior review. Reliability testing was done measuring signal-to-noise ratio, and the results demonstrated high reliability, including when SDS added to risk adjustment.

2b2. Validity Testing

Comments:

**Exclusions appear to be valid and consistant with the evidence presented. None of the groups excluded were done so inappropriately.

The developer provides two risk models:

age and gender composition of the county; and

optional addition to these two variables of the percent of households falling below the federal poverty level

Conceptual rationale for SDS factors ia included included in risk model.

The developer assessed the probability that a county is higher or lower than a benchmark or threshold, given county size—i.e., whether the indicator can discriminate the best performing counties from the lower performing counties. The developer reported:

- The developer reports on discrimination results for counties.
- The developer indicates the "event" is hospitalization for bacterial pneumonia.

The developer reports:

- Identify low performing counties for most counties: 78% of counties can be classified as better or worse than the threshold
- Identify high performing counties for moderate to large counties; 10% of counties can be classified as better or worse than the benchmark.
- The AHRQ QIs use frequently reported administrative data variables. PQI 11 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis.
- Missing data handled – rates are typically less than 1% of the state database.

**No concerns.

**Are the exclusions consistent with the evidence? YES

Are any patients or patient groups inappropriately excluded from the measure? NO, although exclusions reduce numerator by 1%+0.27%+10.5%. Although discharges transferred into a hospital are excluded, these encounters are captured in the area level numerator via the originating hospitalization

Is an appropriate risk-adjustment strategy included in the measure? developer reports the observed to predicated values across the deciles range between 0.99-1.01, indicating it is well calibrated.reports the was 0.58, which it concludes was poor

Were appropriate community factors considered in the conceptual model? likely no, poor c-statistic presumably was due to the limited predictors included

Should race be considered in the risk model? Yes, see disparities section (page 15 of worksheet).

Does this measure identify meaningful differences about quality?

This indicator has strong discrimination to identify low performing counties for most counties;78% of counties can be classified as better or worse than the threshold.

**No serious issues related to validity.

**2b3. Yes, No.

2b4. Yes, but with low discrimination. Would have liked to see whether race was significant. Not sure if they are suggesting that risk adjustment should be part of the measure.

2b5. Low. Concern that only 10% of counties can be classifies as better or worse than benchmark (2nd analysis).

2b6. NA

2b7. Appropriate

**No concerns, missing data handled similarly in these AHRQ measures

** Specifications are somewhat consistent with the evidence. Summary of validity testing is not available from prior review. Only face validity from previous testing was done. Face validity was done through four clinical expert panels. The indicator was overall deemed as "useful" with some specific concerns. No statistical results of face validity testing provided. Overall moderate validity.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:

**It appears that reliability was tested with a large enough patient population to demonstrate whether or not the measure is

appropriate in a larger population. This can not be stated positively ast here is no summary of reliability testing available from the prior review.
The developer indicates there are updates to the reliability testing since the last submission. Reliability testing at the level of the measure score has been conducted using more current data and risk adjustment is included.
Data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) is used: 40 states representing about 89% of U.S. county hospital discharges, for about 30 million hospital discharges from community hospitals. Data were from 2009-2013
The developer provides two risk models:
age and gender composition of the county; and
optional addition to these two variables of the percent of
households falling below the federal poverty level.
**Reliability testing did occur and had an adequate scope. The number of exclusions were higher than expected. The testing was conducted at the score level.
**Do the results demonstrate sufficient reliability so that differences in performance can be identified? Yes, signal-to-noise >0.95.
**The developer indicates there are updates to the reliability testing since the last submission
Reliability testing at the level of the measure score has been conducted using more current data
Risk adjustment is included.
Data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) is used: 40 states representing about 89% of U.S. county hospital discharges, for about 30 million hospital discharges from community hospitals. Data were from 2009-2013
**6, High
**Yes demonstrates using signal to noise, high rating
** Exclusions were noted and consistent with the evidence. No apparent threats identified, including none from missing data. Meaningful differences identified through probability of county being higher or lower than threshold or benchmark. Risk adjustment was done using statistical model and was tested appropriately. Two risk models were provided with 3 variables (age, gender, and poverty level).

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer reports:

- All data elements are in defined fields in electronic claims
- The measure is based on readily available administrative billing and claims data.
- There are no fees. The AHRQ QI software is publically available from the AHRQ Quality Indicators website.
- Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**The developer reports:
- All data elements are in defined fields in electronic claims
- The measure is based on readily available administrative billing and claims data.
- There are no fees. The AHRQ QI software is publically available from the AHRQ Quality Indicators website.
- Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

**The data used to calculate the measure is available in electronic fields in the E.H.R. as well as in administrative data. The measure is feasible and can be put into operational use.
**No concerns, required data are readily available or could be captured without undue burden and can be implemented for performance measurement
**It seems feasible.
All data elements are in defined fields in electronic claims

The measure is based on readily available administrative billing and claims data.
**High
**No concerns
** Data collection obtained through administrative and electronic claims data. No data collection barriers identified. High feasibility.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
The measure is currently being used in the following programs:

- Public Reporting:
  - Arizona Department of Health Services, AZ Hospital Compare
  - Connecticut Department of Health Services, CT Hospital Compare
  - Maine Health Data Organization (MHDO),
  - Nevada Compare Care
  - Oklahoma State Department of Health, MONAHRQ
  - Utah Department of Health, MONAHRQ website
  - Virginia Health Information, MONAHRQ website
  - Washington State, MONAHRQ website
  - California Office of Statewide Health Planning and Development, Healthcare Information Division
  - Connecticut, Office of Health Care Access
  - St John´s Episcopal Hospital
  - Arkansas Department of Human Services: Arkansas Medicaid Performance
  - Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
  - Northwest Hospital and Medical Center
  - Houston and Harris County State of Health Partners
- Payment programs:
  - CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use Reports (QRUR)
- Quality Improvement with benchmarking:
  - West Jefferson Medical Center

Improvement results

- The PQI 11 hospital admissions rate has decreased by 87,000 fewer hospitalizations from 2011-2013
  - In 2012 the rate was 2.7 per 1,000
  - in 2013 the rate was 2.3 per 1,000
- Variation among counties decreased substantially in 2013. The developer is not certain whether this is a single year anomaly or an actual trend.

Unexpected findings (positive or negative) during implementation
The developer notes:

- No challenges implementing this measure. This use of the AHRQ PQIs has grown since the initial endorsement,

suggesting the measure is implementable and useful.

- The transition to ICD-10-CM will provide challenges in understanding time trends or rates from calendar year 2015. This is a challenge all measures based on coded data will encounter.

Potential harms
The developer did not identify potential harms.

Feedback :
No feedback provided on QPS. Measure reviewed by MAP in Value-Based Payment Modifier Program in 2012. MAP recommended to support the direction.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments:
**Current uses of the measure are publicly reported and the measure is currently being used in an accountability program.

Accountability program details
The measure is currently being used in the following programs:

- Public Reporting:
  - Arizona Department of Health Services, AZ Hospital Compare
  - Connecticut Department of Health Services, CT Hospital Compare
  - Maine Health Data Organization (MHDO),
  - Nevada Compare Care
  - Oklahoma State Department of Health, MONAHRQ
  - Utah Department of Health, MONAHRQ website
  - Virginia Health Information, MONAHRQ website
  - Washington State, MONAHRQ website
  - California Office of Statewide Health Planning and Development, Healthcare Information Division
  - Connecticut, Office of Health Care Access
  - St John´s Episcopal Hospital
  - Arkansas Department of Human Services: Arkansas Medicaid Performance
  - Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
  - Northwest Hospital and Medical Center
  - Houston and Harris County State of Health Partners
- Payment programs:
  - CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use Reports (QRUR)
- Quality Improvement with benchmarking:
  - West Jefferson Medical Center
- The PQI 11 hospital admissions rate has decreased by 87,000 fewer hospitalizations from 2011-2013
  - In 2012 the rate was 2.7 per 1,000
  - in 2013 the rate was 2.3 per 1,000
- Variation among counties decreased substantially in 2013. The developer is not certain whether this is a single year anomaly or an actual trend.

The developer has identified no challenges or potential harms in instituting this measure.
**The results of the measure can be used to determine appropriate interventions aimed at a goal of high-quality, efficient healthcare. No unintended consequences were described by the measure developer.
**How can the performance results be used to further the goal of high-quality, efficient healthcare?The rate of PQI 11 hospital admissions has decreased from 2011 to 2013, the variation between counties decreased substantially in 2013;identification of high

rates can help to target populations most in need of intervention. Observation status (with adoption of CMS "2 midnight rule") may impact numerator.
**I don't think thee are significant unintended consequences.
Differences in the measure may lead to investigation of root causes within those measurement groups, including access to care, vaccination rates, other pre-existing conditions, etc. Finding these, and ameliorating them may have very positive impact on health outcomes.
**Moderate usability.
**Reported through AHRQ, no concerns, opportunities for continued QI.
** Measure is currently reported publically through HHS/CMS and multiple state health departments. High usability.

Criterion 5: Related and Competing Measures
Related or competing measures • No related or competing measures were identified.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0279

Measure Title: Bacterial Pneumonia Admission Rate (PQI 11)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1. This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Hospital admissions for Bacterial Pneumonia

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

With access to high quality care, early intervention and appropriate pharmaceutical treatment, this condition can often be managed on an outpatient basis. Access to high quality care and strong community health may minimize the likelihood of milder respiratory conditions progressing to pneumonia, and may also be associated with rapid identification and timely outpatient treatment of pneumonia, to reduce the likelihood of requiring hospitalization.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not applicable.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Please note that this is an outcome measure. Therefore, a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.4.1 and 1a.8 below, to provide additional context and support for the measure.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

The guidelines cited here provide processes to identify and treat community-acquired pneumonia in order to reduce the need for hospitalization. With access to high quality care, early intervention and appropriate pharmaceutical treatment this condition can often be managed on an outpatient basis.

Infectious Disease Society of America (IDSA)
http://www.idsociety.org/Organ_System/

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Not applicable

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).

Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches were performed to update evidence. The most recent evidence review results presented below consist of articles published between January 2012 and October 2015. Additional articles were identified using previous project-related literature reviews. The search terms used included the relevant MeSH term (pneumonia). We combined this clinical search string with (hospital*[Title/Abstract] NOT fungal[Title] NOT viral[Title]) AND (prevent*[Title/Abstract] OR "access to care"[Title/Abstract] OR "ambulatory care sensitive"[Title/Abstract] OR “avoidable hospitalization”[Title/Abstract] OR "small area analysis"[MeSH]). For completeness we also tested more inclusive search strings. Below we have provided a summary of the most up-to-date evidence.

1a.8.2. Provide the citation and summary for each piece of evidence.

Importance

Cost and association with other outcomes

Bacterial pneumonia is a relatively common acute condition, which, if left untreated in a susceptible individual, can lead to death (see pneumonia in-hospital mortality indicator). One large retrospective study of claims data from a health insurance database by Sato et al. (2013) found that mean all-cause total healthcare cost for an inpatient Community-Acquired Pneumonia (CAP) episode ranged from $11,148 to $51,219, depending on risk stratum and age group.[1] Another study by Davydow et al. (2013) analyzed HRS data to show that hospitalization for pneumonia was associated with an average of 1.01 new impairments in Activities and Instrumental Activities of Daily Living (95% CI 0.71-1.32) among patients without baseline functional impairment and an average of 0.99 new impairments in Activities and Instrumental Activities of Daily Living (95% CI 0.57-1.41) among those with mild-to-moderate baseline limitations. Hospitalization for pneumonia was also associated with moderate-to-severe cognitive impairment (OR 2.46; 95% CI 1.60-3.79) and substantial depressive symptoms (OR 1.63; 95% CI 1.06-2.51).[2]

Risk Factors

Studies have also examined patient demographic, behavioral and clinical risk factors for bacterial pneumonia. A study of community-dwelling older adults by Juthani-Mehta et al. (2013) used a bivariate Cox model to show that risk factors associated with pneumonia requiring hospitalization include: age >75 (Hazard ratio 1.37; 95% CI1.03–1.82), history of pneumonia (HR 2.91; 95% CI 1.77–4.78), histamine blocker use (HR 1.38; 95% CI 1.00–1.91), statin use (HR 0.67; 95% CI 0.50–0.90), former smoking status (HR 1.68 95% CI 1.24–2.27), current smoking status (HR 1.95; 95% CI 1.16–3.26), pack years of smoking (HR 1.011; 95% CI 1.007–1.016), BMI loss (HR 1.44; 95% CI 1.01–2.04), unstable BMI (HR 1.66; 95% CI 1.00–2.77) incident mobility limitation (HR 1.68; 95% CI 1.26–2.22), and mean oral plaque score >=1 (HR 1.70; 95% CI 1.27–2.26). However is important to note that more than 50% of the study participants were excluded due to missing data.[3] Finally, a cohort study by Yende et al (2013) found that baseline circulating C-reactive protein (CRP) level was associated with higher risk of pneumonia hospitalization (ORs with 95% CIs per SD increase in CRP for ARIC, CHS, and Health ABC were 1.21 [1.15-1.28], 1.08 [1.02-1.15], and 1.12 [1.02-1.21], respectively, and P values were <0.0001, <0.0001, and <0.04, respectively).[4] In univariate analysis, the study showed the following factors to be associated with a significantly increased risk of pneumonia hospitalization: smoker status, pack-years smoking, BMI <18.5, heart failure, coronary heart disease, diabetes, CKD, and FEV1 <=80, while BMI >25 was shown to be associated with decreased risk of pneumonia hospitalization (all $p < 0.05$).

Validity

As a population health indicator, we consider the relationship of PQI 11 rates to a wide range of factors that are amenable to changes in public policy and community based interventions that may improve access to quality care and community resources, reduce risky personal behaviors or improve self-care, reduce environmental exposure, or prevent the development of pneumonia.

Disparities

Healthy People 2020 defines inequities/disparities in outcomes as "A particular type of health difference that is closely linked with social or economic disadvantage. Health disparities adversely affect groups of people who have systematically experienced greater social or economic obstacles to health based on their racial or ethnic group, religion, socioeconomic status, gender, mental health, cognitive, sensory, or physical disability, sexual orientation, geographic location, or other characteristics historically linked to discrimination or exclusion." Disparities in outcomes among subpopulations highlight the need for improvement in vulnerable populations. Various disparities have been reported in hospitalization rates for bacterial pneumonia. For example, multiple studies have revealed outcome disparities based on gender, race, insurance status and income levels. In one study by AHRQ using a 5% sample of Medicare beneficiaries from 1991 to 1998, directly standardized age-adjusted annual hospital discharge was found to be significantly higher among black men and white men than among women of the same race, respectively (both $p < 0.001$), but among those

of the same gender there was no significant difference found in the rates based on race.[5] Likewise, a study of community-dwelling older adults by Juthani-Mehta et al. (2013) used a bivariate Cox model to show that male gender (HR 2.07; 95% CI 1.54–2.78) is associated with pneumonia requiring hospitalization, however more than 50% of the study participants were excluded due to missing data.[3] A retrospective analysis of low-risk community-acquired pneumonia (CAP) patients discharged from the ED to complete outpatient antibiotic therapy showed that male gender (OR 2.94: 95% CI 1.46-5.89; p = .0018), African American race (OR 0.06; 95% CI 0.31-0.73; p = .004) and having self-pay insurance (OR 0.38; 95% CI 0.19-0.63; p = .0034) were significantly associated with hospitalization among low-risk CAP patients.[6] One recent cohort study by Yende et al (2013) found female gender to be associated with a significantly increased risk of pneumonia hospitalization and black race to be associated with decreased risk of pneumonia hospitalization (all $p < 0.05$),[4] and a Texas cohort study found Latino's rates were 41% higher (p<.0001) than non-Hispanic whites and 90% higher than non Hispanic blacks (p=.0304).[7] A study by Millman et al. reported that among the 11 states studied, pneumonia hospitalizations per 1,000 in population were 5.42 times higher in low-income zip codes than high-income zip codes (statistical significance not provided).[8]

Access to care

The following section discusses evidence related to leverage points within the health care system.

We only identified one study that examined pneumonia hospitalization rates as a function of provider density. A Texas cohort study evaluated PQI 11 county rates and county characteristics. They found an increase of 1 primary care physician per 1000 population is associated with a decrease in age-sex-ethnicity adjusted PQI 11 rates by 33% (p=.0158)[7]

This indicator is subject to seasonal variation and variation due to infectious outbreaks (e.g. influenza epidemic). Influenza vaccination and pneumococcal vaccination reduced hospitalization and ED visits for pneumonia/influenza at a patient level in several studies.[9-18] In a recent study of inpatient discharge records from 500 non-federal short-stay U.S. hospitals by Simonsen et al. (2014) used a negative binomial multiple regression model to determine that 13-valent pneumococcal conjugate vaccine was associated with significant reductions in hospital admissions for all-cause pneumonia for one subset of adults—those age 18-39 (12% [95% CI 6–17]—however reductions for other adult age groups were not significant. The vaccine also significantly reduced admissions (by 25-37% depending on age group and type of pneumonia) for non-invasive pneumococcal or lobar pneumonia in all adults age groups and for invasive pneumococcal pneumonia for adults age 18-39 and those >=65 of age (all p <0.05), indicating herd protection.[17] In another study, following implementation of a Veterans Affairs performance measurement program targeted at increasing vaccination for pneumonia, vaccination rates increased. [18] During this time, pneumonia hospitalization rates decreased by 50% among elderly Veterans Health Administration enrollees but increased among Medicare enrollees by 15% (P for differences in trend <.001). Similar reductions in admissions after vaccination have been found in Spain[19] and Japan[20] but was not demonstrated in one Candian study[21].

Risk models can be used to target high-risk populations to prevent infection or promote early intervention. One cohort study by Yende et al (2013) developed a simple clinical risk prediction model to predict 10-year risk of pneumonia hospitalization in a derivation cohort using logistic regression. The authors also determined the clinical application of the model in the external validation cohort by estimating the number of pneumonia cases identified using different risk cutoffs. Most subjects had a predicted risk > 5%, and since all participants in the external validation cohort were > 65 years of age, the CDC guidelines for pneumococcal vaccination would require targeting all subjects for prevention. Targeting those with predicted risk >5%, >10%, and >15% predicted risk would target 90.4%, 39.1%, and 14.4% of the subjects in this cohort and identify 88.9%, 55.1%, and 26.8% of pneumonia cases, respectively. To illustrate a strategy of targeting a more select population, such as those with risks of less than or equal to 10% in the clinical risk prediction model, would require treating approximately one-third of the population and identify one-half of pneumonia cases. Targeting only high-risk participants (>= 15% risk) or 14% of the population would identify one-fourth of pneumonia cases, a subgroup that could be targeted initially to conduct randomized clinical trials to test pharmacologic interventions cost-effectively. The risk prediction model had moderate discrimination and excellent calibration (AUC=0.77; 95% CI 0.76-0.79; HL C statistic, 0.12). Model discrimination was similar in the internal validation cohort (AUC=0.77; HL C statistic, 0.65). Model discrimination was lower in external validation cohort (AUC= 0.62), but calibration was excellent (HL C statistic, 0.45).[4]

Variation in Admission Practice Patterns

While some patients will develop pneumonia that clearly requires hospitalization, the decision to treat patients on an inpatient basis may be based on local practice patterns. We found few studies that specifically examined the impact of practice patterns and admission decisions in a geographic area on pneumonia hospitalization rates. However, two studies examined variation in hospitalization rates. Using a nationally representative sample of 5,899 nursing home residents in 815 facilities, residents with suspected pneumonia in not-for-profit facilities were found to be hospitalized at a lower rate (14.6 percent) than residents in for-profit facilities (29.4 percent), while those in government facilities were hospitalized more frequently (41.1 percent) than those in for profit facilities (both $p<0.01$).[22] In constrast, another French study examined the effect of using the pneumonia severity index (PSI) as a decision aid for admitting practices related to community-acquired pneumonia in 16 emergency departments. Those emergency departments utilizing the severity index as an admitting decision aid did not have subsequent hospitalization and ICU admission rates that were significantly different from those of non-PSI user emergency departments.[23]

References

1. Sato R, Gomez Rey G, Nelson S, Pinsky B. Community-acquired pneumonia episode costs by age and risk in commercially insured US adults aged >/=50 years. *Applied health economics and health policy.* 2013;11(3):251-258.
2. Davydow DS, Hough CL, Levine DA, Langa KM, Iwashyna TJ. Functional disability, cognitive impairment, and depression after hospitalization for pneumonia. *The American journal of medicine.* 2013;126(7):615-624.e615.
3. Juthani-Mehta M, De Rekeneire N, Allore H, et al. Modifiable risk factors for pneumonia requiring hospitalization of community-dwelling older adults: the Health, Aging, and Body Composition Study. *Journal of the American Geriatrics Society.* 2013;61(7):1111-1118.
4. Yende S, Alvarez K, Loehr L, et al. Epidemiology and long-term clinical and biologic risk factors for pneumonia in community-dwelling older Americans: analysis of three cohorts. *Chest.* 2013;144(3):1008-1017.
5. Baine WB, Yu W, Summe JP. Epidemiologic trends in the hospitalization of elderly Medicare patients for pneumonia, 1991-1998. *Am J Public Health.* 2001;91(7):1121-1123.
6. Aliyu ZY, Aliyu MH, McCormick K. Determinants for hospitalization in " low-risk" community acquired pneumonia. *BMC infectious diseases.* 2003;3(1):11.
7. Lemus FC, Tan A, Eschbach K, Freeman Jr DH, Freeman JL. Correlates of Bacterial Pneumonia Hospitalizations in Elders, Texas Border. *Journal of immigrant and minority health.* 2010;12(4):423-432.
8. Millman M, ed *Committee on Monitoring Access to Personal Health Care Services.* Washington, D.C.: National Academy Press; 1993. Acess to health care in America/ Committee on Monitoring Access to Personal Health Care Services, Institute of Medicine.
9. Christenson B, Lundbergh P. Comparison between cohorts vaccinated and unvaccinated against influenza and pneumococcal infection. *Epidemiol Infect.* 2002;129(3):515-524.
10. Hak E, Nordin J, Wei F, et al. Influence of high-risk medical conditions on the effectiveness of influenza vaccination among elderly members of 3 large managed-care organizations. *Clin Infect Dis.* 2002;35(4):370-377.
11. Davis JW, Lee E, Taira DA, Chung RS. Influenza vaccination, hospitalizations, and costs among members of a Medicare managed care plan. *Med Care.* 2001;39(12):1273-1280.
12. Nordin J, Mullooly J, Poblete S, et al. Influenza vaccine effectiveness in preventing hospitalizations and deaths in persons 65 years or older in Minnesota, New York, and Oregon: data from 3 health plans. *J Infect Dis.* 2001;184(6):665-670.
13. Crocetti E, Arniani S, Bordoni F, Maciocco G, Zappa M, Buiatti E. Effectiveness of influenza vaccination in the elderly in a community in Italy. *Eur J Epidemiol.* 2001;17(2):163-168.
14. Christenson B, Hedlund J, Lundbergh P, Ortqvist A. Additive preventive effect of influenza and pneumococcal vaccines in elderly persons. *Eur Respir J.* 2004;23(3):363-368.
15. Herzog NS, Bratzler DW, Houck PM, et al. Effects of previous influenza vaccination on subsequent readmission and mortality in elderly patients hospitalized with pneumonia. *Am J Med.* 2003;115(6):454-461.
16. Puig-Barbera J, Diez-Domingo J, Perez Hoyos S, Belenguer Varea A, Gonzalez Vidal D. Effectiveness of the MF59-adjuvanted influenza vaccine in preventing emergency admissions for pneumonia in the elderly over 64 years of age. *Vaccine.* 2004;23(3):283-289.

17. Simonsen L, Taylor RJ, Schuck-Paim C, Lustig R, Haber M, Klugman KP. Effect of 13-valent pneumococcal conjugate vaccine on admissions to hospital 2 years after its introduction in the USA: a time series analysis. *The Lancet. Respiratory medicine.* 2014;2(5):387-394.
18. Jha AK, Wright SM, Perlin JB. Performance measures, vaccinations, and pneumonia rates among high-risk patients in Veterans Administration health care. *Am J Public Health.* 2007;97(12):2167-2172.
19. Domínguez A, Izquierdo C, Salleras L, et al. Effectiveness of the pneumococcal polysaccharide vaccine in preventing pneumonia in the elderly. *European Respiratory Journal.* 2010;36(3):608-614.
20. Kawakami K, Ohkusa Y, Kuroki R, et al. Effectiveness of pneumococcal polysaccharide vaccine against pneumonia and cost analysis for the elderly who receive seasonal influenza vaccine in Japan. *Vaccine.* 2010;28(43):7063-7069.
21. Johnstone J, Eurich DT, Minhas JK, Marrie TJ, Majumdar SR. Impact of the pneumococcal vaccine on long-term morbidity and mortality of adults at high risk for pneumonia. *Clinical infectious diseases.* 2010;51(1):15-22.
22. Konetzka RT, Spector W, Shaffer T. Effects of nursing home ownership type and resident payer source on hospitalization for suspected pneumonia. *Medical care.* 2004;42(10):1001-1008.
23. Renaud B, Coma E, Labarere J, et al. Routine use of the Pneumonia Severity Index for guiding the site-of-treatment decision of patients with pneumonia in the emergency department: a multicenter, prospective, observational, controlled cohort study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2007;44(1):41-49.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PQI11_NQF_0279_Measure_Evidence_Form_151214.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This indicator is intended to identify hospitalizations for pneumonia, either specified as bacterial or unspecified organism. With access to high quality care, early intervention and appropriate pharmaceutical treatment this condition can often be managed on an outpatient basis.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental files.

Table 1. Reference Population Rate and Distribution of County Performance for PQI 11
Overall Reference Population Rate

Year	Number of Counties (Numerator)a	Number of Events (Denominator)a	Population at Risk	Observed Rate Per 1,000a
2009	3,137	702,634	232,379,612	3.0236
2010	3,141	668,530	234,909,365	2.8459

2011	3,144	678,908	237,419,828	2.8595
2012	3,141	638,091	237,830,861	2.683
2013	3,140	550,294	240,482,275	2.2883

Distribution of County-level Observed Rates in Reference Population Per 1,000

Year	Number of Counties(p=percentile)b	Mean	SD	p5	p25	Median	p75	p95
2009	3,137	5.20	25.36	0.05	2.53	3.91	5.83	10.92
2010	3,141	5.26	30.93	0.05	2.38	3.78	5.57	10.06
2011	3,144	5.24	30.91	0.08	2.40	3.73	5.55	10.11
2012	3,141	4.59	30.89	0.10	2.03	3.27	4.74	8.25
2013	3,140	3.28	2.43	0.02	1.68	2.99	4.48	7.59

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

aThe observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all counties included in the reference population data (denominator). Note: Observations from counties with rates outside of 1.5*interquartile range are excluded as outliers.

bThe distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile. Note: Counties with rates outside of 1.5*interquartile range are excluded as outliers.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

n/a

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

This table is also included in the supplemental files.

Table 2. Admission Rates per 1,000 (PQI 11), by patient and hospital characteristics, 2013

Patient/hospital characteristic	Estimate	Std Error	p-value	(Ref Grp = *)	Lower 95% CL	Upper 95% CL
Total U.S.	228.46	0.3068		227.85	229.06	
Patient Characteristics						
Age Groups:						
18-39*	65.75	0.6612		64.45	67.04	
40-64	184.14	0.5186	<.001	183.12	185.15	
65 and over	346.51	0.4670	<.001	345.60	347.43	
Gender:						
Male*	241.55	0.4640		240.64	242.46	
Female	218.28	0.4090	<.001	217.48	219.08	
Patient Zip Code Median Income						
First quartile (lowest income)		329.72	1.0955	<.001	327.57	331.86
Second quartile	280.84	0.7294	<.001	279.41	282.27	
Third quartile	238.88	0.6304	<.001	237.65	240.12	
Fourth quartile (highest income)*		189.72	0.4304		188.87	190.56
Location of patient residence (NCHS):						
Rural	352.65	2.1849	<.001	348.37	356.94	
Urban*	225.95	0.3098		225.35	226.56	
Location of Care:						
Northeast*	182.813	0.701		181.44	184.19	
Midwest	285.255	0.652	<.001	283.98	286.53	
South	242.848	0.507	<.001	241.86	243.84	
West	187.330	0.657	<.001	186.04	188.62	

Source: Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project, Nationwide Inpatient Sample, 2013, and AHRQ Quality Indicators, version 6.0.
Rates are adjusted by age and gender using the AHRQ QI PQI Reference Population for 2013 as the standard population; when reporting is by age, the adjustment is by gender only; when reporting is by gender, the adjustment is by age only.
NCHS - National Center for Health Statistics designation for urban-rural locations.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
n/a

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Bacterial pneumonia is a relatively common acute condition, which, if left untreated in a susceptible individual, can lead to death (see pneumonia in-hospital mortality indicator). One large retrospective study of claims data from a health insurance database by Sato et al. (2013) found that mean all-cause total healthcare cost for an inpatient Community Acquired PNE (CAP) episode ranged from $11,148 to $51,219, depending on risk stratum and age group. (Statistical significance not indicated).(1) Another study by Davydow et al. (2013) analyzed HRS data to show that hospitalization for pneumonia was associated with 1.01 new impairments in Activities and Instrumental Activities of Daily Living (95% CI 0.71-1.32) among patients without baseline functional impairment and 0.99 new impairments in Activities and Instrumental Activities of Daily Living (95% CI 0.57-1.41) among those with mild-to-moderate baseline limitations. Hospitalization for pneumonia was also associated with moderate-to-severe cognitive impairment (OR 2.46; 95% CI 1.60-3.79) and substantial depressive symptoms (OR 1.63; 95% CI 1.06-2.51).(2)

In 2011, pneumonia was the second most common principal diagnosis recorded in US hospital stays, accounting for approximately 3,818,000 stays.(3) Table 1 shows that in the AHRQ QI 2013 PQI Reference Population there were 550,294 qualifying discharges with a rate of 2.3 per 1,000.4 The coefficient of variation was 0.74 in 2013.(4)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Sato R, Gomez Rey G, Nelson S, Pinsky B. Community-acquired pneumonia episode costs by age and risk in commercially insured US adults aged >/=50 years. Applied health economics and health policy. 2013;11(3):251-258.
2. Davydow DS, Hough CL, Levine DA, Langa KM, Iwashyna TJ. Functional disability, cognitive impairment, and depression after hospitalization for pneumonia. The American journal of medicine. 2013;126(7):615-624.e615.
3. Pfuntner A, Wier LM, Stocks C. Most Frequent Conditions in U.S. Hospitals, 2011. HCUP Statistical Brief #162. September 2013. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb148.pdf. Accessed August 19, 2013.
4. HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
n/a

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Prevention, Pulmonary/Critical Care, Pulmonary/Critical Care : Pneumonia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://1.usa.gov/1OHkljs Note: The URL link currently provides Version 5.0 specifications. Version 6.0 specifications will be released publicly March 2016 and found via the module page: http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PQI11_Technical_Specifications_v6.1alpha_151214_v02.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
As standard protocol, the AHRQ QI program annually updates all measures with Fiscal Year coding changes, refinements based on stakeholder input, refinements to improve specificity and sensitivity based on additional analyses, and necessary software changes. In addition, approximately every two years, AHRQ updates the risk adjustment parameter estimates and composite weights based on the most recent year of data (i.e., the most current reference population possible). The refined measures are tested and confirmed to be valid and reliable prior to release of the updated software.

Since the last update, the following changes have been made to the indicator:
- Added ICD-9-CM diagnosis codes 452.40 Staphylococcal pneumonia NOS, 482.41 Methicillin susceptible staphylococcal pneumonia, 482.42 MRSA pneumonia, 482.49 Staphylococcal pneumonia NEC to numerator definition of pneumonia
- Remove numerator exclusions: MDC14 (pregnancy, childbirth, and puerperium); MDC15 (newborn and other neonates). By definition, discharges with a principal diagnosis of bacterial pneumonia are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QITM software does not explicitly exclude obstetric cases.
- Added exclusion of patients with any diagnosis code or procedure code for Immunocompromised state
- The data upon which to base the reference population was updated. Updated with 2013 US Census population estimates
- Fiscal Year coding updates

Additional information regarding revisions to PQI software and technical specifications available online:
http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx
Note: Version 6.0 specifications will be released publicly March 2016.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Discharges, for patients ages 18 years and older, with a principal ICD-9-CM or ICD-10-CM-PCS diagnosis code for bacterial pneumonia.

[NOTE: By definition, discharges with a principal diagnosis of bacterial pneumonia are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QI software does not explicitly exclude obstetric cases.]

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Users may specify a time period; but the time period is generally one year. Note that the reference population rates and signal variance parameters assume a one-year time period.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Please see attached excel file in S.2b. for Version 6.0 specifications.
Prevention Quality Indicators technical specifications and appendices also available online at http://www.qualityindicators.ahrq.gov/Modules/PQI_TechSpec.aspx). Note: The URL link currently provides Version 5.0 specifications. Version 6.0 specifications will be released publicly March 2016.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Population ages 18 years and older in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The term "metropolitan area" (MA) was adopted by the U.S. Census in 1990 and referred collectively to metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs), and primary metropolitan statistical areas (PMSAs). In addition, "area" could refer to either 1) FIPS county, 2) modified FIPS county, 3) 1999 OMB Metropolitan Statistical Area, or 4) 2003 OMB Metropolitan Statistical Area. Micropolitan Statistical Areas are not used in the QI software.

See AHRQ QI website for 2014 Population File Denominator report for calculation of population estimates embedded within AHRQ QI software programs. http://www.qualityindicators.ahrq.gov/Downloads/Software/SAS/V50/AHRQ_QI_Population_File_V50.pdf

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Not applicable.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Not applicable.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific*

Acceptability)

The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups). An option model is available that includes percent of households under the federal poverty level as well. Because we cannot individually observe the age and gender of each person in a counties population, we use the age and gender distribution of the county to estimate the number of "cases" in each age*gender group. The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Data (SID) for the year 2013 (combined), a database consisting of 40 states, and the U.S. Census data by county. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., area). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Additional information on methodology can be found in the Empirical Methods document on the AHRQ Quality Indicator website (www.qualityindicators.ahrq.gov) and in the supplemental information attached.

The specific covariates for this measure are as follows:

PARAMETER	LABEL
SEX	Female
AGE	Male, Age 18-24
AGE	Male, Age 25-29
AGE	Male, Age 30-34
AGE	Male, Age 35-39
AGE	Male, Age 40-44
AGE	Male, Age 45-49
AGE	Male, Age 50-54
AGE	Male, Age 55-59
AGE	Male, Age 60-64
AGE	Male, Age 65-69
AGE	Male, Age 70-74
AGE	Male, Age 75-79
AGE	Male, Age 80-84
AGE	Male, Age 85+
AGE	Female, Age 18-24
AGE	Female, Age 25-29
AGE	Female, Age 30-34
AGE	Female, Age 35-39
AGE	Female, Age 40-44
AGE	Female, Age 45-49
AGE	Female, Age 50-54
AGE	Female, Age 55-59
AGE	Female, Age 60-64
AGE	Female, Age 65-69
AGE	Female, Age 70-74
AGE	Female, Age 75-79
AGE	Female, Age 80-84
AGE	Female, Age 85+
POVCAT	Poverty Decile 2
POVCAT	Poverty Decile 3
POVCAT	Poverty Decile 4
POVCAT	Poverty Decile 5
POVCAT	Poverty Decile 6
POVCAT	Poverty Decile 7
POVCAT	Poverty Decile 8
POVCAT	Poverty Decile 9
POVCAT	Poverty Decile 10 (Highest percent poverty)1

1Deciles are based on the percentage of households under the federal poverty level (FPL).
Source: http://qualityindicators.ahrq.gov/Modules/pqi_resources.aspx
Parameter estimates with and without SES covariates (POVCAT) are included with the Technical Specifications.
Please note Version 6.0 will be released publicly in March 2016.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 observed rate of each PQI is simply the number of individuals living in a county admitted to the hospital for the condition of interest divided by the census population estimate for the area (adult population for adult measures and child population for pediatric measures). The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected performance observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic distributions observed in the user's dataset? The expected rate is calculated only for risk-adjusted indicators.

The expected rate is estimated for each county using logistic regression.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the performance observed in the user's dataset were applied to a mix of patients with demographics distributed like the reference population? The risk adjusted rate is calculated using the indirect method as observed rate divided by expected rate multiplied by the reference population rate. The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural counties).

For additional information, please see supporting information in the Quality Indicator Empirical Methods attached in the supplemental files.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing), or county (PSTCO=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications and software are specified to be used with any ICD-9-CM- or ICD-10-CM/PCS coded administrative billing/claims/discharge dataset.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Population : County or City

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: All community based care

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
n/a

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PQI11_NQF_0279_Measure_Testing_Form_151214v02.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0279
Measure Title: Bacterial Pneumonia Admission Rate (PQI 11)
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set***

of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.

- **For <u>all</u> measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

<u>Note</u>: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in**

performance;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record

☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2009-2013. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ)[1]. HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 40 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

The SID data elements include ICD-9-CM coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

1.3. What are the dates of the data used in testing?
HCUP data: 2009-2013.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice

[1] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Population health	☒ other: County

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1. Reference Population Rate and Distribution of County Performance, for PQI 11 Bacterial Pneumonia Admission Rate

Overall Reference Population Rate				
Year	**Number of Counties**	**Outcome of Interest (Numerator)[1]**	**Population at Risk (Denominator)[1]**	**Observed Rate Per 1000[1]**
2009	3,137	702,634	232,379,612	3.0236
2010	3,141	668,530	234,909,365	2.8459
2011	3,144	678,908	237,419,828	2.8595
2012	3,141	638,091	237,830,861	2.6830
2013	3,140	550,294	240,482,275	2.2883

Distribution of County-level Observed Rates in Reference Population Per 1000								
Year	**Number of Counties**	**(p=percentile)[2]**						
		Mean	**SD**	**p5**	**p25**	**Median**	**p75**	**p95**
2009	3,137	5.20	25.36	0.05	2.53	3.91	5.83	10.92
2010	3,141	5.26	30.93	0.05	2.38	3.78	5.57	10.06
2011	3,144	5.24	30.91	0.08	2.40	3.73	5.55	10.11
2012	3,141	4.59	30.89	0.10	2.03	3.27	4.74	8.25
2013	3,140	3.28	2.43	0.02	1.68	2.99	4.48	7.59

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population for the county in the reference population data (denominator). For PQI 11 this includes population ages 18 years and older.

[2]The distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

See 1.5 (Table 1)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We use SDS variables in risk adjustment. Because the measures are applied at the county level, we use variables that describe the make-up of the county population. Two risk models are available; one includes age and sex make-up of the county, the other also includes the percent of households falling below the federal poverty level. These data are obtained from the US Census.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Signal-to-noise

The signal-to-noise ratio refers to the entire population of US counties, comparing the degree to which rates are different from county to county (the signal) to how stable the rates are within counties (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance between counties, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences between counties (e.g. one county performs better than others).

The signal-to-noise ratio is estimated for each county. The overall signal-to-noise estimate is an average of county-level signal to noise ratios weighted by county size. County size is calculated as the eligible population for PQI 11 (population 18 years and older). Weighting by county size reduces the impact of counties that have very small denominators (the number of patients at risk).

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one county's performance from the other counties in the population, it is sensitive to the distribution of county sizes as well as the distribution of observed rates in the reference population. If the counties in a population all have performance in a narrow range, it is more difficult to reliably distinguish between counties' performance than when county performance is spread out over a much wider range. For example, if all counties have nearly perfect performance, it will be impossible to distinguish between them. As a consequence, if the distribution of county rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal to noise ratio. Different methods of calculating reliability and signal-to-noise result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8. To account for the uncertainty (noise) in a county's performance due to reliability concerns stemming from low volume, smoothed rates can be

calculated.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a. Signal-to-Noise Ratio by Size Decile, for PQI 11 Bacterial Pneumonia Admission Rate

Size Decile	Number of Counties	Avg. Number of Qualifying Population per County in Decile	Avg. Signal-to-Noise Ratio for Counties in Decile
1	314	3152.4	0.85651
2	314	6721.4	0.92496
3	314	9936.5	0.95092
4	314	13659	0.96401
5	314	18291.9	0.97324
6	314	24976.6	0.98030
7	314	34729.2	0.98559
8	314	53886.1	0.99060
9	314	103898.5	0.99483
10	314	496615.5	0.99829
Overall	3140	76586.7	0.96566

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 2b. SES Signal-to-Noise Ratio by Size Decile, for PQI 11 Bacterial Pneumonia Admission Rate

Size Decile	Number of Counties	Avg. Number of Qualifying Population per County in Decile	Avg. Signal-to-Noise Ratio for Counties in Decile
1	314	3152.4	0.83927
2	314	6721.4	0.91365
3	314	9936.5	0.94308
4	314	13659	0.95812
5	314	18291.9	0.96879
6	314	24976.6	0.977
7	314	34729.2	0.98315
8	314	53886.1	0.989
9	314	103898.5	0.99395
10	314	496615.5	0.99799
Overall	3140	76586.7	0.96113

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The indicator demonstrates good reliability with a signal-to-noise ratio of 0.97. Reliability remains strong for all county sizes. Smoothed rates, which are recommended for all counties (and are implemented in the AHRQ software), address

any remaining reliability concerns for the smallest counties. When SES is added to the risk adjustment, the reliability remains high at 0.96.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Systematic Assessment of Face Validity

In 2008-2009, we convened clinical panels to assess the use of the PQI across a wide range of applications, one of which was comparative reporting of area level rates.[2] We solicited nominations from national professional organization to form four clinical expert groups for a total of 73 panelists. We utilized a hybrid approach for the panel review, using two review processes, which were conducted simultaneously with information exchange between the two panels and a third and fourth panel that conducted specific reviews based on specialty. The development of this hybrid process builds from the experiences in previous panel evaluations of QI modules. The panel process that has been employed during the development of the PSIs, the PDIs and the validation of the IQIs is based on the RAND-UCLA Appropriateness Method and is termed a "nominal group" panel. The approach allowed for a wider range of input is fully described elsewhere.[3]

Panelists rated the indicator on appropriateness of use after completing a 14 item questionnaire. The questionnaire evaluated the face validity of the indicators, the panelists' perspectives on bias and potential for gaming, and the overall usefulness of the indicators when applied at one of three aspects of the health care system: area, payer and large provider organizations, for one of three purposes: internal quality improvement, comparative reporting (either public or not), and pay for performance.

Support was defined as follows:

- Full support for use: Median score of 7-9 without disagreement
- Some concern regarding use: Median score of 4-6.9 regardless of agreement status
- General support with some concerns regarding use due to disagreement: Median score of 7-9 with disagreement
- Major concern regarding use: Median score or 1-3.9 regardless of agreement status

---

[2] Davies SM, McDonald KM, Schmidt E, Schultz E, Geppert J, Romano, PS. 'Expanding the Uses of AHRQ's Prevention Quality Indicators: Validity from the Clinician Perspective'. Medical Care. 2011 Aut;49(8):479-85.

[3] Davies S, Romano PS, Schmidt EM, Schultz E, Geppert JJ, McDonald KM. Assessment of a Novel Hybrid Delphi and Nominal Groups Technique to Evaluate Quality Indicators. *Health services research.* 2011;46(6 Pt 1):2005-2018.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Face validity: Clinical Panel Review

During the clinical panel review both the Delphi and Nominal panel supported the measure with some concern. Other topics discussed by the panel included:

- Panelists cited access to vaccinations as a crucial aspect of preventing pneumonia and subsequent hospitalization.
- Panelists felt that this indicator reflects access to care more than quality of care.
- Patient factors may limit the control the healthcare system has over admission rates. These factors include comorbidities, socioeconomic status, geographic limitations (transportation issues), propensity to present in a timely manner, and cultural differences or beliefs.
- Comorbidities (COPD/Asthma, diabetes, HIV) and patient self-care (smoking) need careful consideration as risk covariates for some applications
- Use and affordability of antibiotics may be enhanced with coverage from payer organizations; however, panelists also expressed concern for antibiotic overuse and the emergence of antibiotic-resistant strains in some populations (e.g. long-term care).
- The panelists felt this indicator may also reflect some amount of "social" hospital admissions. In other words, cases in which the physician determines social support or the home environment are insufficient for recovery outside of the hospital.
- This indicator may aid in exposing geographic areas that may benefit from increased targeting of resources such as vaccinations.
- The Bacterial Pneumonia indicator may be pertinent in long-term care.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

The clinical panel noted that this indicator overall was useful. They identified specific actions that could improve rates, especially from a population health perspective, such as access to medications, reduction of risk factors, such as COPD/asthma, diabetes and smoking and increasing vaccination in vulnerable populations. However, it is important, as with many population health indicators to acknowledge the complex factors influencing the indicators.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Using the 2013 data from 40 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

A total of 1783 discharges were excluded due to diagnoses of sickle cell disease. Removing this exclusion would increase the numerator count by 0.27%. A total of 69,135 discharges were excluded due to diagnoses of

immunocompromised state. Removing this exclusion would increase the numerator count by 10.5%. The denominator does not change. Although discharges transferred into a hospital are excluded, these encounters are captured in the area level numerator via the originating hospitalization.

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The exclusion for sickle cell has been retained to increase the face validity of the measure and to align the measure with the pediatric PDI for pneumonia admissions, although the impact on the numerator is minor for PQI 11. Patients meeting the exclusion are identifiable without additional burden using the same data as is used to identify numerator qualifying discharges.

Pneumonias in special populations with immunocompromised states are excluded since infections in these populations may be more likely to progress or require hospitalization despite access to high quality care.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 3 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Sociodemographic Factors

The risk model includes age and gender of the population as pneumonia prevalence and severity increase with age.

We considered multiple income risk factors for inclusion in the model, including % of households under the federal poverty level, household income, median % of poverty level (e.g. 200% of federal poverty level). The Public Health Disparities Geocoding Project has completed extensive evaluation of alternative income variables and has demonstrated that the percent poverty variable consistently detects expected gradients in health across health outcomes, is widely

available, has a low rate of missing data even at the census tract level[4]. Our team has explored alternative income variables that do not outperform the poverty level variable (data not shown).

We also considered a conceptual model that acknowledged the impact of community factors such as clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise along with community norms and beliefs can impact prevention of pneumonia and self-care, which in turn can impact hospitalization rates. Poverty can be a mitigating factor, inasmuch as impoverished communities are more likely to experience housing and food insecurity[5], air pollution[6], occupational exposure and have higher smoking rates. The relationship between environment, income and health is complex and the mechanism is not fully understood.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The process to select risk factors is described in the AHRQ QI Empirical Methods report. The results of the analyses are provided in the PQI Parameter Estimates document. Both documents are available to reviewers in the supporting materials. The results of the analyses are provided in the tables below as well as on the submitted excel spreadsheet.

There are several steps involved in estimating the QI risk-adjustment models.

1. Construct candidate covariates
2. Select model covariates
3. Estimate the models
4. Evaluate the models

Covariates are coded for each discharge record based on the data elements, data values, and logic described in the technical specifications and the appendices of the risk-adjustment coefficient tables. For a given covariate, if the discharge meets the technical specification for that covariate a value of "1" is assigned to the discharge level covariate data element. Otherwise a value of "0" is assigned to the discharge level covariate data element.

[4] Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian S. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures-the public health disparities geocoding project. *American journal of public health.* 2003;93(10):1655-1671.

[5] Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the US. American journal of preventive medicine. 2009;36(1):74-81. e10.

[6] Bell ML, Ebisu K. Environmental inequality in exposures to airborne particulate matter components in the United States. *Environmental health perspectives.* 2012;120(2):1699-1704.

Table 3a. Risk Adjustment Coefficients, for PQI 11 Bacterial Pneumonia Admission Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCEPT		1	-3.749527	0.0046693	644812.26	<.0001
SEX	Female	1	-0.284201	0.0060377	2215.6369	<.0001
AGE	Male, Age 18-24	1	-4.595648	0.0174480	69374.252	<.0001
AGE	Male, Age 25-29	1	-4.391109	0.0183105	57510.547	<.0001
AGE	Male, Age 30-34	1	-4.078057	0.0160250	64759.791	<.0001
AGE	Male, Age 35-39	1	-3.833633	0.0149193	66026.858	<.0001
AGE	Male, Age 40-44	1	-3.539106	0.0127695	76812.656	<.0001
AGE	Male, Age 45-49	1	-3.189231	0.010956	84722.552	<.0001
AGE	Male, Age 50-54	1	-2.839348	0.0093576	92066.203	<.0001
AGE	Male, Age 55-59	1	-2.53314	0.0086037	86684.869	<.0001
AGE	Male, Age 60-64	1	-2.242185	0.0082550	73773.335	<.0001
AGE	Male, Age 65-69	1	-1.849783	0.0078224	55919.063	<.0001
AGE	Male, Age 70-74	1	-1.371769	0.0075174	33298.206	<.0001
AGE	Male, Age 75-79	1	-0.935044	0.0073683	16103.794	<.0001
AGE	Male, Age 80-84	1	-0.515653	0.0072582	5047.2643	<.0001
AGE	Male, Age 85+		Referent	.	.	.
AGE	Female, Age 18-24	1	0.356899	0.0243836	214.23652	<.0001
AGE	Female, Age 25-29	1	0.468113	0.0248619	354.51293	<.0001
AGE	Female, Age 30-34	1	0.430415	0.0218146	389.29451	<.0001
AGE	Female, Age 35-39	1	0.476901	0.0200486	565.83003	<.0001
AGE	Female, Age 40-44	1	0.477837	0.0171045	780.43034	<.0001
AGE	Female, Age 45-49	1	0.455280	0.0146924	960.21994	<.0001
AGE	Female, Age 50-54	1	0.407625	0.0125792	1050.0632	<.0001
AGE	Female, Age 55-59	1	0.316078	0.0116759	732.83162	<.0001
AGE	Female, Age 60-64	1	0.255288	0.0112498	514.95605	<.0001
AGE	Female, Age 65-69	1	0.219237	0.0106560	423.29124	<.0001
AGE	Female, Age 70-74	1	0.151179	0.0102428	217.84426	<.0001
AGE	Female, Age 75-79	1	0.085940	0.0100073	73.749014	<.0001
AGE	Female, Age 80-84	1	0.021800	0.0097536	4.9955119	0.0254
AGE	Female, Age 85+		Referent			
c-statistic=0.5793						

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 3b. SES Risk Adjustment Coefficients, for PQI 11 Bacterial Pneumonia Admission Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCEPT		1	-3.94948999	0.006499255	369279.291	<.0001
SEX	Female	1	-0.28457968	0.006038849	2220.74923	<.0001
AGE	Male, Age 18-24	1	-4.61345947	0.01744986	69898.7685	<.0001
AGE	Male, Age 25-29	1	-4.40304578	0.018311667	57816.3456	<.0001
AGE	Male, Age 30-34	1	-4.08651070	0.016026102	65020.3727	<.0001
AGE	Male, Age 35-39	1	-3.83857848	0.014920245	66189.4800	<.0001
AGE	Male, Age 40-44	1	-3.54038971	0.012770469	76857.8832	<.0001
AGE	Male, Age 45-49	1	-3.18848882	0.010957868	84667.6775	<.0001
AGE	Male, Age 50-54	1	-2.83836855	0.009358752	91981.7573	<.0001
AGE	Male, Age 55-59	1	-2.53402316	0.008604836	86723.3417	<.0001
AGE	Male, Age 60-64	1	-2.24543402	0.008256171	73967.8960	<.0001
AGE	Male, Age 65-69	1	-1.85405674	0.007823491	56162.2957	<.0001
AGE	Male, Age 70-74	1	-1.37765528	0.007518665	33573.7327	<.0001
AGE	Male, Age 75-79	1	-0.94197166	0.007369664	16337.2995	<.0001
AGE	Male, Age 80-84	1	-0.51953681	0.007259365	5121.95086	<.0001
AGE	Male, Age 85+		Referent	.	.	.
AGE	Female, Age 18-24	1	0.35642016	0.024383944	213.656747	<.0001
AGE	Female, Age 25-29	1	0.46790745	0.024862224	354.193021	<.0001
AGE	Female, Age 30-34	1	0.43133784	0.021815029	390.952099	<.0001
AGE	Female, Age 35-39	1	0.47801132	0.020049045	568.445682	<.0001
AGE	Female, Age 40-44	1	0.47888730	0.017105029	783.824873	<.0001
AGE	Female, Age 45-49	1	0.45618517	0.014692954	963.971259	<.0001
AGE	Female, Age 50-54	1	0.40727556	0.012579804	1048.16338	<.0001
AGE	Female, Age 55-59	1	0.31494004	0.011676639	727.478377	<.0001
AGE	Female, Age 60-64	1	0.25433035	0.011250562	511.031835	<.0001
AGE	Female, Age 65-69	1	0.21852429	0.010656798	420.480574	<.0001
AGE	Female, Age 70-74	1	0.15029291	0.010243706	215.259739	<.0001
AGE	Female, Age 75-79	1	0.08499939	0.010008377	72.1280732	<.0001
AGE	Female, Age 80-84	1	0.02054758	0.009754873	4.43688629	0.0352
AGE	Female, Age 85+		Referent	.	.	.
POVCAT	Poverty Decile 2	1	0.01932578	0.006627971	8.50184359	0.0035
POVCAT	Poverty Decile 3	1	0.12198983	0.006446426	358.104124	<.0001
POVCAT	Poverty Decile 4	1	0.19068058	0.006298865	916.406896	<.0001
POVCAT	Poverty Decile 5	1	0.22932931	0.006233634	1353.43225	<.0001
POVCAT	Poverty Decile 6	1	0.21239576	0.00630114	1136.1964	<.0001
POVCAT	Poverty Decile 7	1	0.15093824	0.006434241	550.305620	<.0001
POVCAT	Poverty Decile 8	1	0.30588576	0.006301617	2356.21304	<.0001
POVCAT	Poverty Decile 9	1	0.36064149	0.006274725	3303.40658	<.0001
POVCAT	Poverty Decile 10 (Highest percent poverty)[1]	1	0.40956802	0.006214897	4342.94467	<.0001
c-statistic=0.5786						

[1]Deciles are based on the percentage of households under the federal poverty level (FPL).

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

SES variables are consistent across PQI risk adjustment.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest (i.e. admission for bacterial pneumonia). The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic. The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model based on the value of the observations covariates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. This creates a set of pairs of observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the non-event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common "goodness of fit" tests because these tests tend to not be informative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Table 4a. Age-sex Risk adjustment Model Discrimination and Calibration, for PQI 11 Bacterial Pneumonia Admission Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	44,318,065	0.000267	0.000265
2	33,916,474	0.000429	0.00043
3	31,395,312	0.000677	0.000676
4	30,224,021	0.001100	0.001132
5	29,401,580	0.001664	0.001671
6	25,602,365	0.002282	0.002254
7	18,215,356	0.003756	0.003698
8	12,187,383	0.006309	0.006364
9	8,074,508	0.010539	0.010473
10	7,133,225	0.018453	0.018526
C-Statistic	0.5794		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 4b. Age-sex and SES Risk adjustment Model Discrimination and Calibration, for PQI 11 Bacterial Pneumonia Admission Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	45,858,273	0.000265	0.000265
2	35,102,508	0.000438	0.000428
3	30,298,903	0.000693	0.000681
4	31,705,487	0.001125	0.001108
5	29,456,027	0.001700	0.001687
6	23,604,733	0.002406	0.002464
7	17,537,714	0.003789	0.003814
8	12,115,605	0.006489	0.006452
9	7,758,448	0.010620	0.010860
10	7,030,591	0.018744	0.018519
C-Statistic	0.5787		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

See Table 4 in 2b4.6

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

See Table 4 in 2b4.6

2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e.*, *what do the results mean and what are the norms for the test conducted*)

A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated, as the observed to predicted values across the deciles range between 0.99– 1.01. The discrimination is poor with a c-statistic of 0.58, presumably due to the limited predictors included. Addition of SES to the model results in no change in model performance.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

This analysis assesses the probability that a county is higher or lower than a benchmark or threshold, given county size. It reflects whether the indicator can discriminate the best performing counties from the lower performing counties.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of counties have a lower admission rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance).

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across counties of various sizes. Each county is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each county the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Counties are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each county to determine if the county rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 5 reports the proportion of counties above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of counties not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 5a. Performance Categories by County Size Decile for PQI 11 Bacterial Pneumonia Admission Rate

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Discharges Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	314	3,152.40	0.0000	0.9045	0.0955	0.5414	0.0796	0.379
2	314	6,721.40	0.0000	0.8694	0.1306	0.4522	0.2325	0.3153
3	314	9,936.50	0.0000	0.8503	0.1497	0.4809	0.2452	0.2739
4	314	13,659.00	0.0000	0.8471	0.1529	0.4745	0.2325	0.293
5	314	18,291.90	0.0000	0.8694	0.1306	0.4873	0.2834	0.2293
6	314	24,976.60	0.0000	0.9045	0.0955	0.5064	0.2771	0.2165
7	314	34,729.20	0.0032	0.8694	0.1274	0.5605	0.242	0.1975
8	314	53,886.10	0.0701	0.8854	0.0445	0.6306	0.1847	0.1847
9	314	103,898.50	0.1146	0.8567	0.0287	0.7771	0.1146	0.1083
10	314	496,615.50	0.0924	0.8981	0.0095	0.949	0.0191	0.0319
Overall	3,140	76,586.70	0.028	0.8755	0.0965	0.586	0.1911	0.2229

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 5b. SES Performance Categories by County Size Decile for PQI 11 Bacterial Pneumonia Admission Rate

	Benchmark	Threshold

Size Decile	Number of Counties	Average Number of Denominator Discharges Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	314	3,152.40	0.0000	0.8567	0.1433	0.5159	0.0860	0.1401
2	314	6,721.40	0.0000	0.8439	0.1561	0.4427	0.2420	0.1401
3	314	9,936.50	0.0000	0.8503	0.1497	0.4618	0.2484	0.1369
4	314	13,659.00	0.0000	0.8408	0.1592	0.4650	0.2389	0.1115
5	314	18,291.90	0.0000	0.8662	0.1338	0.4713	0.2930	0.1592
6	314	24,976.60	0.0032	0.8981	0.0987	0.4936	0.2898	0.1210
7	314	34,729.20	0.0255	0.8662	0.1083	0.5478	0.2516	0.0764
8	314	53,886.10	0.0955	0.8822	0.0223	0.6115	0.2134	0.0828
9	314	103,898.50	0.1242	0.8567	0.0191	0.7643	0.1242	0.0637
10	314	496,615.50	0.0987	0.8949	0.0064	0.9459	0.0287	0.0191
Overall	3,140	76,586.70	0.0347	0.8656	0.0997	0.5720	0.2016	0.1051

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e.*, *what do the results mean in terms of statistical and meaningful differences?*)

This indicator has strong discrimination to identify low performing counties for most counties; 78% of counties can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold). The indicator has strong discrimination, particularly for moderate to large counties to identify high performing counties; 10% of counties can be classified as better or worse than the benchmark. Performance discrimination remains strong when adding SES to risk adjustment.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what*

statistical analysis was used)
Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PQI 11 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all county discharge records. The rate of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

For these variables, rates of missing data are typically less than 1% of the state database. It is unlikely the bias would occur from such a low rate of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases for missing data is appropriate.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

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

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because the indicator is based on readily available administrative billing and claims data and U.S. Census data, feasibility is not an issue.
The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees. Software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting

	Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2011/Methodology.html Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings Maine Health Data Organization (MHDO), MONAHRQ Website http://gateway.maine.gov/mhdo/monahrq/Methodology.html Nevada Compare Care, MONAHRQ website http://nevadacomparecare.net/ Oklahoma State Department of Health, MONAHRQ https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html Utah Department of Health, MONAHRQ website https://health.utah.gov/myhealthcare/monahrq/ Virginia Health Information, MONAHRQ website http://www.vhi.org/MONAHRQ/default.asp?yr=2013 Washington State, MONAHRQ website http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions California Office of Statewide Health Planning and Development, Healthcare Information Division http://oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/ Connecticut, Office of Health Care Access http://www.ct.gov/dph/lib/dph/ohca/publications/2010/prev_hosp_report01-2010.pdf Houston and Harris County State of Health Partners http://houstonstateofhealth.org/soh_doc/ Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW) http://www.healthindicators.gov/Resources/Initiatives/CMS/Prevention-Quality-Indicators-Report_20/Indicator/Report Northwest Hospital and Medical Center http://www.nwhospital.org/downloads/pdfs/Northwest-Hospital-CHNA-2013.pdf Payment Program CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use Reports (QRUR) http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2011-ACSC-Outcomes-Measures.pdf Quality Improvement with Benchmarking (external benchmarking to multiple organizations) West Jefferson Medical Center http://www.wjmc.org/docs/WJMC-Secondary-Data-Profile-09-23-2013.pdf

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting:
Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona.
http://pub.azdhs.gov/hospital-discharge-stats/2011/Methodology.html

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Connecticut.
http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine.
http://gateway.maine.gov/mhdo/monahrq/Methodology.html

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada: Quality reporting on hospitals across the state of Nevada Under NV Regulation R151-8 this transparency website presents hospital quality and utilization information.
http://nevadacomparecare.net/

Oklahoma State Department of Health, MONAHRQ
Compares quality ratings on hospitals across Oklahoma.
https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html

Utah Department of Health, MONAHRQ website
Hospital quality ratings from all hospitals in Utah.
https://health.utah.gov/myhealthcare/monahrq/

Virginia Health Information, MONAHRQ website
Compares quality ratings on hospitals across Virginia.
http://www.vhi.org/MONAHRQ/default.asp?yr=2013

Washington State, MONAHRQ website
Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals.
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions

California Office of Statewide Health Planning and Development, Healthcare Information Division
OSHPD Patient Discharge Data from all hospitals in California, totaling over 4 million records annually.
http://oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/

Connecticut, Office of Health Care Access
Preventable Hospitalizations in Connecticut: A Current Assessment of Access to Community Health Services: 2004-2009 state- and county-level hospital admission rate data from most hospitals in CT.
http://www.ct.gov/dph/lib/dph/ohca/publications/2010/prev_hosp_report01-2010.pdf

Houston and Harris County State of Health Partners
State of Health Annual Report patterned after HCPHES Annual Report.
http://houstonstateofhealth.org/soh_doc/

Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
Purpose of the HIW is to: Provide a single, user-friendly, source for national, state, and community health indicators; Facilitate harmonization of indicators across initiatives; Link indicators with evidence-based interventions.
http://www.healthindicators.gov/Resources/Initiatives/CMS/Prevention-Quality-Indicators-Report_20/Indicator/Report

Northwest Hospital and Medical Center
Community Health Needs Assessment (CHNA) required by federal law and Patient Protection and Affordable Health Care Act.
http://www.nwhospital.org/downloads/pdfs/Northwest-Hospital-CHNA-2013.pdf

Quality Improvement:
West Jefferson Medical Center
Reports indicators of potentially avoidable hospitalizations associated with the parish in which it is located, and compared those indicators with state-level indicators.
http://www.wjmc.org/docs/WJMC-Secondary-Data-Profile-09-23-2013.pdf

Payment Programs:
CMS Medicare FFS Physician Feedback Program/Value-Based Payment Modifiers and Quality and Resource Use Reports (QRUR)1
Program includes measures of Ambulatory Care Sensitive Conditions (ACSC), used by Physicians receiving Medicare FFS payment

modifiers.
http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2011-ACSC-Outcomes-Measures.pdf
1The numerator of PQI 11 is used in this program. The denominator and risk adjustment is modified to meet specific program requirements.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
n/a

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
n/a

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

See Table 1 in response to question 1b.2.

The rate of PQI 11 hospital admissions has decreased from 2011 to 2013. In 2012 the rate was 2.7 per 1,000 while in 2013 the rate was 2.3 per 1,000. This decrease represents over 87,000 fewer hospitalizations. Further, the variation between counties decreased substantially in 2013, although we cannot determine whether this is a single-year anomaly or an actual trend. Additionally, it is important to note, the standard deviation is highly sensitive to outliers and the observed change appeared to be due to outlier counties.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
During a structured clinical panel review, panelists postulated that some uses of this indicator could disincentive care for high risk individuals. However, no evidence of this unintended consequence has arisen during actual use of the indicator. Rather, identification of high rates can help to target populations most in need of intervention.
Panelists in the same structured review and subsequent expert panel review noted that treatment of bacterial pneumonia in observation care may substitute for inpatient treatment, that this substitution may be systematic between areas and that this will impact the rate of the indicator. During a literature review, we identified no studies that specifically examined observation stays as a substitute for inpatient care. In a retrospective analysis of a 2002-2011 large administrative claims database of commercially insured individuals in the USA, pneumonia was one of the most common short inpatient stay diagnoses that may be impacted by the so-called CMS “2 midnight rule”.1 A retrospective analysis of observation stays from three distinct data source: 2010 Atlanta hospitals protocol driven observation units, 2010 Georgia hospitals for observation units (including protocol-driven, discretionary care and all

bed locations), and 2009-10 National Hospital Ambulatory Medical Care Survey (NHAMCS) for similarly diverse of observation units found that pneumonia was the 10th and 11th most common diagnosis in two of the three study settings.2

1. Overman RA, Freburger JK, Assimon MM, Li X, Brookhart MA. Observation stays in administrative claims databases: underestimation of hospitalized cases. Pharmacoepidemiology and drug safety. Sep 2014;23(9):902-910.
2. Ross MA, Hockenberry JM, Mutter R, Barrett M, Wheatley M, Pitts SR. Protocol-driven emergency department observation units offer savings, shorter stays, and reduced admissions. Health Aff (Millwood). Dec 2013;32(12):2149-2156.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PQI11_NQF0279_Supplemental_Files_151214.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Carol, Stocks, Carol.Stocks@ahrq.hhs.gov
Co.3 **Measure Developer if different from Measure Steward:**
Co.4 **Point of Contact:**

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The following panelists participated in a 2009 structured panel review of the Agency for Healthcare Research and Quality Prevention Quality Indicators, which focused on evaluating expansion of the indicators to alternative denominator populations. The panel used a modified Delphi approach to evaluate the indicators, using a method that combined a nominal group technique and a Delphi technique.1 All panelists rated the indicators and received feedback from other panelists. The panelists participated in a conference call to discuss the indicators and the discussion was summarized and distributed to the group before final rating. Some panelists requested that their affiliation with this report remain anonymous, and this list is therefore a partial representation of the individuals that comprised the panels in their entirety.

1. Davies S, McDonald KM, Schmidt E, Geppert J, Romano PS. Expanding the uses of AHRQ's Prevention Quality Indicators: Validity from the clinician perspective. Med Care. Aug 2011; 49(8): 679-685.

Sandra G. Adams, MD, MS, FCCP
Pulmonolary & Critical Care Medicine
South Texas Veterans Health Care System
University of Texas Health Science Center
San Antonio, Texas
Nominated by American College of Chest Physicians

Wilbert S. Aronow, MD, FACC, FAHA, AGSF, FCCP
Geriatric Medicine, Cardiology
Westchester Medical Center
New York Medical College
Valhalla, New York
Nominated by The American Geriatrics Society

John D. Birkmeyer, MD, FACS
Surgery
University of Michigan Hospital
Ann Arbor, Michigan
Nominated by American College of Surgeons

James H. Black, III, MD
Vascular Surgery
Johns Hopkins Hospital
Johns Hopkings University
Baltimore, Maryland
Nominated by Society for Vascular Surgery

Cynthia Boyd, MD
Geriatric Medicine
Johns Hopkins Hospital, Bayview Medical Center
Johns Hopkins University
Batlimore, Maryland
Nominated by The American Geriatrics Society

Suzanne Bradley, MD

Geriatric Medicine and Infectious Diseases
Veterns Affairs Ann Arbor Healthcare System
University of Michigan
Ann Arbor, Michigan
Nominated by The American Geriatrics Society

Vickie Brown, RN, MPH, CIC
Hospital Epidemiology
University of North Carolina Hospital
Chapel Hill, North Carolina
Nominated by Association for Professionals in Infection Control and Epidemiology

John Buckley, MD, MPH
Pulmonary & Critical Care Medicine
Henry Ford Hospital
Wayne State University
Detroit, Michigan
Nominated by Society of General Internal Medicine

Pauline M. Camacho, MD, FACE
Division of Endocrinology and Metabolism
Loyola University Medical Center
Loyola Stritch School of Medicine
Maywood/Chicago, Illinois
Nominated by American Academy of Clinical Endocrinologists

Dickson Cheung, MD, MBA, MPH
Emergency Medicine
Sky Ridge Medical Center
Lone Tree, Colorado
Nominated by American College of Emergency Physicians

Leslie Davis, MSN, RN, ANP-C
Division of Cardiology
Department of Medicine
University of North Carolina
Chapel Hill, North Carolina
Nominated by American Academy of Nurse Practitioners

Barbara DeBaun, MSN, RN, CIC
Improvement Advisor
Beacon, Bay Area Patient Safety Collaborative
Bay Area Counties, California
Nominated by Association for Professionals in Infection Control and Epidemiology

Gregory J. Dehmer, MD, FACC, FSCAI
Interventional Cardiology
Scott & White Healthcare
Texas A&M University Health Science Center
Temple, Texas
Nominated by American College of Cardiology

Shawkat Dhanani, MD, MPH
Geriatric Medicine
Veterans Affairs Greater Los Angeles Healthcare System
University of California at Los Angeles
Los Angeles, California

Nominated by The American Geriatrics Society

Michelle Farber, RN, CIC
Infection Prevention and Control
Mercy Community Hospital
Coon Rapids, Minnesota
Nominated by Association for Professionals in Infection Control and Epidemiology

Amy Fendrich, MD
Internal Medicine
Memorial Regional Hospital Primary Care Clinics
Dania Beach, Florida
Nominated by American Puclic Health Association, Medical Care Section

Carlos M. Ferrario, MD, FAHA, FASA, FACC
Cardiology
Wake Forest University Health Science Center
Winston-Salem, North Carolina
Nominated by American College of Cardiology

John E. Gardella, MD, MBA, FCCP, FHM
Hospital Medicine
Presbyterian Hospital, Novant Health
Charlotte, North Carolina
Nominated by Society of Hospital Medicine

Eric Gertner, MD, MPH, FACP
Internal Medicine
Lehigh Valley Hospital
Penn State College of Medicine
Allentown, Pennsylvania
Nominated by American College of Physicians

James M. Gill, MD, MPH
Family Medicine
Delaware Valley Outcomes Research
Jefferson Medical College
Wilmington, Delaware
Nominated by American Puclic Health Association, Medical Care Section

Louis Gilleran, MD, MPH, FACPM
Preventive Aerospace Medicine
Naval Medical Center San Diego
San Diego, California
Nominated by American Puclic Health Association, Medical Care Section

Christopher Gonzalez, MD, MBA
Urology
Northwestern University Medical Center
Chicago, Ilinois
Nominated by American Urological Association

Michael K. Gould, MD, MS
Pulmonary and Critical Care Medicine
Veterans Affairs Palo Alto Health Care System
Stanford University
Palo Alto, California

Nominated by American Thoracic Society

George Grunberger, MD, FACP, FACE
Diabetes, Endocrinology
Grunberger Diabetes Institute
Wayne State University
Bloomfield Hills & Detroit, Michigan
Nominated by American Academy of Clinical Endocrinologists

A. Seiji Hayashi, MD, MPH
Family Medicine
Unity Health Care, Inc.
George Washington University
Washinton, DC
Nominated by American Puclic Health Association, Medical Care Section

Richard Hellman, MD, FACP, FACE
Clinical Endocrinology
North Kansas City Hospital
University of Missouri, Kansas City School of Medicine
Kansas City, Missouri
Nominated by American Academy of Clinical Endocrinologists

Steve J. Hodges, MD
Surgery, Pediatric Urology
Wake Forest University Health Sciences
Institute for Regenerative Medicine
Winston-Salem, North Carolina
Nominated by American College of Surgeons (by proxy through Dr. Anthony Atala)

Mary Johnson, MS, RD, CDE, BC-ADM
Diabetes Quality and Education
Geisinger Health System
Central Pennsylvania
Nominated by American Dietetic Association

Jeanette Kalupa, MSN, ACNP-BC, APNP
Acute Care Nurse Practioner, Hospitalist
Cogent Healthcare
Aurora St. Luke´s Medical Center
Milwaukee, Wisconsin
Nominated by American Academy of Nurse Practitioners

Marjorie L. King, MD, FACC, FAACVPR
Cardiology
Helen Hayes Hospital
Columbia University
West Haverstraw, New York
Nominated by American College of Cardiology

Jerry A. Krishnan, MD, PhD
Pulmonary and Critical Care Medicine
University of Chicago, Asthma and COPD Center
Chicago, Ilinois
Nominated by American Thoracic Society

Geoffrey Lamb, MD

Internal Medicine
Froedtert Hospital
Medical College of Wisconsin
Milwaukee, Wisconsin
Nominated by Society of General Internal Medicine

Gene Lambert, MD, MBA
Hospital Medicine
Massachusetts General Hospital
Harvard University
Boston, Massachusetts
Nominated by Society of Hospital Medicine

David E. Lanfear, MD, MS, FACC
Cardiology
Henry Ford Hospital
Wayne State University
Detroit, Michigan
Nominated by American College of Cardiology

Karen Lasser, MD, MPH
Primary Care, Internal Medcine
Boston University, Medical Center
Boston, Massachusetts
Nominated by Society of General Internal Medicine

Susan Lee, MS, NP-C, FAANP
Family Medicine, Internal Medicine
Bloomington Hospital
Bloomington, Indiana
Nominated by American Academy of Nurse Practitioners

Edgar V. Lerma, MD
Nephrology
Associates in Nephrology, S.C.
University of Illinois at Chicago
Chicago, Illinois
Nominated by The American Society of Nephrology

James T. Li, MD, PhD
Internal Medicine, Allergy
Mayo Clinic, College of Medicine
Rochester, Minnesota
Nominated by American Academy of Allergy Asthma and Immunology

Jenny J. Lin, MD
Internal Medicine
Mount Sinai Medical Center
Mount Sinai Medical School
New York, New York
Nominated by Society of General Internal Medicine

Frank LoGerfo, MD
Vascular Surgery
Beth Israel Deaconess Hospital
Harvard Medical School
Boston, Massachusetts

Nominated by Society for Vascular Surgery

John J. Lopez, MD, FACC
Interventional Cardiology
Loyola University Medical Center
Stritch School of Medicine, Loyola University
Maywood, Illinois
Nominated by American College of Cardiology

Thomas D. MacKenzie, MD, MSPH
Internal Medicine
Denver Health
University of Colorado
Denver, Colorado
Nominated by Society of General Internal Medicine

William Marston, MD
Vascular Surgery
University of North Carolina
Chapel Hill, North Carolina
Nominated by Society for Vascular Surgery

Helmut Meisl, MD, FACEP
Emergency Medicine
Good Samaritan Hospital
San Jose, California
Nominated by American College of Emergency Physicians

Mark L. Metersky, MD, FCCP
Pulmonary and Critical Care Medicine
University of Connecticut, Health Center
Farmington, Connecticut
Nominated by American College of Chest Physicians

Lena M. Napolitano, MD, FACS, FCCM
Surgery
University of Michigan
Ann Arbor, Michigan
Nominated by American College of Surgeons

John A. Parker, Jr., MD
Family Medicine
Marshall University
Huntington, West Virginia
Nominated by American Academy of Family Physicians

Jay I. Peters, MD, FCCP
Chief Pulmonary & Critical Care Medicine
University of Texas Health Science Center
San Antonio, Texas
Nominated by American College of Chest Physicians

Michael P. Phelan, MD, FACEP
Emergency Medicine
Cleveland Clinic
Cleveland, Ohio
Nominated by American College of Emergency Physicians

Mark Potter, MD
Family Medicine
University of Illinois at Chicago, Medical Center
Chicago, Illinios
Nominated by American Academy of Family Physicians

Carol Rasmussen, MSN, NP-C, CDE
Diabetes Program
Exodus Healthcare
Magna, Utah
Nominated by American Association of Diabetes Educators

H. David Reines, MD, FACS
Surgery
Inova Fairfax Hospital
Virginia Commonwealth University
Falls Church, Virginia
Nominated by American College of Surgeons

Barry Saver, MD, MPH
Family Medicine
University of Massachusetts Memorial
Worcester, Massachusetts
Nominated by American Academy of Family Physicians

Douglas J.E. Schuerer, MD, FACS, FCCM
Acute and Critical Care Surgery
Barnes Jewish Hospital
Washington University
St. Louis, Missouri
Nominated by American College of Surgeons

June Schulz, RRT, FAACVPR
Respiratory Care, Pulmonary Rehabilitation
Sanford University of South Dakota Medical Center
University of South Dakota
Sioux Falls, South Dakota
Nominated by American Association of Cardiovascular and Pulmonary Rehabilitation

Kristine M. Thompson, MD
Emergency Medicine
Mayo Clinic
Jacksonville, Florida
Nominated by American College of Emergency Physicians

Francesca J. Torriani, MD, FIDSA
Infectious Disease, Epidemiology
University of California at San Diego
San Diego, California
Nominated by Infectious Disease Association of California

Dace L. Trence, MD, FACE
Division of Metabolism, Endocrinology, and Nutrition
University of Washington Medical Center
Seattle, Washington
Nominated by American Academy of Clinical Endocrinologists

Arjun K. Venkatesh, MD, MBA
Emergency Medicine
Brigham and Women´s Hospital, Massachusetts General Hospital
Harvard University
Boston, Massachusetts
Nominated by American College of Emergency Physicians

Raoul Wolf, MBBCh, FCCP, FAAAI
Allergy and Clinical Immunology
La Rabida Chidrens Hospital, Comer
University of Chicago
Chicago, Illinois
Nominated by American College of Chest Physicians

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2007

Ad.3 **Month and Year of most recent revision:** 11, 2007

Ad.4 **What is your frequency for review/update of this measure?** Annually

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0283
Measure Title: Asthma in Younger Adults Admission Rate (PQI 15)
Measure Steward: Agency for Healthcare Research and Quality
Brief Description of Measure: Admissions for a principal diagnosis of asthma per 1,000 population, ages 18 to 39 years. Excludes admissions with an indication of cystic fibrosis or anomalies of the respiratory system, obstetric admissions, and transfers from other institutions.
Developer Rationale: This indicator is intended to identify hospitalizations for asthma in younger adults age 18-39. With appropriate pharmaceutical and other outpatient management, risk of hospitalization is decreased.

Numerator Statement: Discharges, for patients ages 18 through 39 years, with a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for asthma.

[NOTE: By definition, discharges with a principal diagnosis of asthma are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QI software does not explicitly exclude obstetric cases.]
Denominator Statement: Population ages 18 through 39 years in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.
Denominator Exclusions: Not applicable.

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Population : County or City

IF Endorsement Maintenance – Original Endorsement Date: Nov 15, 2007 **Most Recent Endorsement Date:** Nov 15, 2007

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

The developer notes:

- This is a population-level measure. The level of analysis is county or city.

- New evidence is provided since the last endorsement maintenance review.
- A systematic review of the body of evidence is not required for outcome measures
    - The developer provides the following rationale: The measure is intended to identify hospitalizations for asthma in younger adults age 18-39. Appropriate pharmaceutical and other outpatient management will decrease risk of hospitalization.
- Although not required per NQF guidance, the developer conducted a literature review (January 2012-October 2015) related to aspects of hospitalization for asthma, as follows:
    - Disparities
    - Environmental exposure
    - *Access to care*
    - Seasonal variation, influenza and vaccination

Questions for the Committee:

- *Although the developer provides updated evidence related to aspects of hospitalization for pneumonia, does the Committee agree the underlying rationale for the measure remains reasonable and there is no need for repeat discussion and vote on Evidence?*
- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer reports:

- Asthma is one of the most common chronic lower respiratory diseases. It is a leading cause of morbidity/ mortality and has patient/societal consequences of poor quality.
- In 2012, an estimated 18.7 million adults had asthma. Asthma results in 1.8 million emergency department visits in 2011 and was responsible for 3,311 deaths in 2012 (age-adjusted rate of 1.0 per standard population of 100,000)."

Reference Population Rate and Distribution of County Performance for PQI 15

Overall Reference Population Rate

Year	Number of Counties	Number of Events (Numerator)	Population at Risk (Denominator)	Observed Rate Per 1,000
2009	3,135	52,986	91,475,217	0.5792
2010	3,138	46,476	91,767,953	0.5065
2011	3,141	43,685	92,184,336	0.4739
2012	3,139	43,746	90,798,464	0.4818
2013	3,140	34,549	91,667,214	0.3769

Distribution of County-level Observed Rates in Reference Population Per 1,000

(p=percentile)

Year	Number of Counties	Mean	SD	p5	p25	Median	p75	p95
2009	3,135	0.50	0.63	0.00	0.00	0.38	0.70	1.51
2010	3,138	0.46	0.63	0.00	0.00	0.34	0.64	1.38
2011	3,141	0.40	0.53	0.00	0.00	0.29	0.56	1.25
2012	3,139	0.67	17.85	0.00	0.00	0.26	0.51	1.07

2013	3,140	0.28	0.37	0.00	0.00	0.18	0.42	0.90

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Disparities

The developer notes:

- Disparities in asthma hospitalization rates among non-Whites, particularly among Blacks, has been documented in numerous studies. Using 2000-2010 National Inpatient Sample (NIS) and HCUP State Inpatient Databases (SID) data, AHRQ documented gender, racial, and income-based disparities for this measure.
    - In 2010, females had a 129% higher rate of hospital stays than males (163.0 versus 71.2 hospital stays per 100,000 population)
    - African American and Hispanic patients had higher rates of asthma hospitalization (297.9 and 144.6 per 100,000 population, respectively) than White and Asian and Pacific Islander patients (90.5 and 65.4 per 100,000 population, respectively)
    - Adult patients in the lowest income communities had higher rates of hospital stays for asthma than those in the highest income communities (194.3 versus 72.6 hospital stays per 100,000 population).

Question for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**yes, "Asthma among young adults age 18-39, for the most part, can be treated in the outpatient setting. Numerous studies have shown an association at the patient level, between appropriate treatment and hospital admission rates"

**This is a health outcome measure. Asthma among young adults age 18-39, for the most part, can be treated in the outpatient setting. Numerous studies have shown an association at the patient level, between appropriate treatment and hospital admission rates, although education interventions have not always had an impact on admission rates.1,2 Asthma admission rates are also associated conceptually, and in some cases empirically, with community-level pollution3 and smoking rates.

**Outcomes measure where the evidence does relate.

**The measure is for an outcome. The rationale is strong that it would be impacted by a variety of things which are mentioned, including access to care, access to interventions such as pharmacologic therapy and various environmental factors.

**Outcome measure. I accept that there is a relationship in any patient between appropriate care and the risk of hospitalization in a given patient. I wish that there was more evidence of improvements in quality correlating with improvements in this measure since it's been in place.

**Agree that there is a relationship between health actions (or inactions) and hospital admissions. Evidence continues to support looking at incidence of hospitaliations for asthma so that hospitals can look at types of interventions to provide to decrease those numbers

1b. Performance Gap

Comments:

**Tables provide Reference Population Rate and Distribution of County Performance provided, Admission Rates by age, gender, Patient Zip Code Median Income,Location of patient residence (NCHS), and location of care provided.
Tables so not report race/ethnicity in data but developer provides lit review regarding in hospitalizations rates among non-Whites, particularly among Blacks.
**Yes.
Asthma affects large numbers and is a leading cause of morbidity/mortality. It has high resource use and Patient/societal consequences are of poor quality. The evidence presented shows disparities data from the measure as specified (current and over time) by population group e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.
**The information provided demonstrated a performance gap in gender, racial, and income-based disparities for this measure,
**There do seem to be gaps that are demonstrated by the measure.
Disparities have also been documented including for gender, race =AA, and based pn community income level.
**Yes, high.

**Data provided highlights disparate hospitaliations for females, African Americans, Hispanics, and low income patients. Targeted interventions to those groups could positively impact hospitaliation rates for asthmatics

1c. High Priority (previously referred to as High Impact)

Comments:

**statement that this issues affects large numbers, is a leading cause of morbidity/mortality, requires high resource use, has patient/societal consequences of poor quality

**n/a

**NA

**N/a

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Administrative claims
-

Specifications:

- The developer indicates some changes to the measure specifications since the last endorsement review:
  - Added exclusion for cases with any listed ICD-9-CM diagnosis codes for cystic fibrosis and anomalies of the respiratory system.
  - Changed eligible age range for numerator and denominator to age 18-39 following clinical panel review. Asthma in older adults (ages 40 and above) is captured in PQI 05.
  - Removed exclusion for discharges in MDC 14 and MDC 15. By definition, discharges with a principal diagnosis of asthma are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QITM software does not explicitly exclude obstetric cases.
  - The data upon which to base the reference population was updated with 2013 U.S. Census population estimates
  - Fiscal Year coding updates (none in this time period)
- The numerator of this measure is: *Number of discharges, for patients ages 18 through 39 years, with a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for asthma.*
- The denominator of this measure is: *Population ages 18 through 39 years in metropolitan area or county.* Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.
- This outcome measure is risk adjusted, using a statistical risk model.
- The calculation algorithm is stated in S.18.
- The measure specifications and software are specified to be used with any ICD-9-CM or ICD-10-CM/PCS coded administrative billing/claims/discharge dataset.

Question for the Committee:

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*

2a2. Reliability Testing Testing attachment

Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Summary of reliability testing is not available from the prior review.

Describe any updates to testing

- The developer indicates there are updates to the reliability testing since the last submission
  - Reliability testing at the level of the measure score has been conducted using more current data.
  - Data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) is used:
    - 40 states representing about 89% of U.S. county hospital discharges, for a total of about 30 million hospital discharges from community hospitals.
    - The HCUP dataset included data from 2009-2013; CHR data from 2014; and ACS data from 2013.
  - The developers created a county characteristic dataset, merged with county-level observed and risk adjusted rates from the HCUP dataset. Candidate predictor variables from the County Health Rankings (CHR) dataset and the American Community Survey (ACS) were analyzed
- The developer provides two risk models:
  - age and gender composition of the county; and
  - optional addition to these two variables of the percent of households falling below the federal poverty level

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Reliability testing was conducted at the performance measure level, but not the individual element level.
- Testing was conducted using signal-to-noise analysis assessing the reliability to confidently distinguish the performance among counties.
- Specifically, the signal-to-noise ratio refers to the entire population of U.S. counties, comparing the degree to which rates are different from county to county (the signal) to how stable the rates are within counties (the noise).

Results of reliability testing

- The developer reported a signal-to-noise ratio of 0.75, which the developer states indicates strong reliability. The developer notes, however, reliability does not meet threshold for the smallest counties with eligible populations under approximately 3,800 individuals.
- The developer reported that when SDS is added to the risk adjustment, the signal-to-noise ratio is 0.74.

Guidance from the Reliability Algorithm : 1 → 2 → 4 → 5 → 6 (highest eligible rating is HIGH)

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Summary of validity testing is not available from the prior review.

Describe any updates to validity testing:

- Empirical validity testing of the measure score was conducted.
- The developer states it sought to "assess the relationship of county-level hospital admission rate for COPD [SIC—NQF staff question whether this should be asthma] with county level measures of socioeconomic status (SES) and community environment, heath behaviors and individual risk factors (i.e. smoking, physical activity and obesity), and access to quality care measures (i.e. primary care physician and other primary care provider density, diabetes screening testing, uninsurance)."

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- The measure was tested for face validity with input from four clinical expert panels involving 73 panelists; the panel was convened from 2008-2009.
- The measure was tested for empirical validity. Developers assessed the relationship of county-level hospital admission rate with county level measures of SES and community environment, heath behaviors and individual risk factors and access to quality care measures.

Validity testing results:

- The developer reports the panels indicated the measure was useful. Specific actions could improve rates, such as access to medications, patient education, reduction of risk factors, such as environmental exposure to pollution or allergens and smoking. There are complex factors influencing the measure.
- For empirical testing, the developer reports:
  - Prevalence, health behaviors (HB) and SES/Environment were statistically significant predictors ($p<.0001$) (see Table 3a).
  - Access to care (AC) was not significant when HB and SES/E are included in the model.
  - Categorizing the variables into interpretable groups (HB, AC, and SES/E) resulted in significant collinearity in the models. In particular, there was a correlation of magnitude 0.77 between IHB and one of the components for SES/E and correlations of magnitude between 0.40 and 0.50 between the components for AC and the other two components for SES/E (see Table 3b). Hence the relative importance of those factors should be interpreted with caution.
- The developer concludes SES explained the most variance, but that moderate correlation between SES status and other factors suggest it is difficult to fully ascertain the unique contribution of each factor on asthma hospitalizations.
- The developer further notes that its disparities analysis found zip codes in the highest income quartile have 31% lower admission rates than those in the lowest income quartile. The developer states that, from a population health perspective, such disparities argue for the importance of the indicator in capturing poor outcomes for

vulnerable populations and that by taking a population health perspective, efforts to decrease individual risk factors, such as obesity, smoking, and limited physical exercise (the variables included in the health behaviors factor) may decrease prevalence and exacerbations of asthma.

Question for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer reports the following:

- Excludes cases with missing gender, age, quarter, year, principal diagnosis or county. Per the developer, these exclusions never exceed 1% of eligible records.
- 262 discharges were excluded due to diagnoses of cystic fibrosis and anomalies of the respiratory system. Removing this exclusion would increase the numerator count by 0.13%. The exclusion of cystic fibrosis and anomalies of the respiratory system has been retained to increase the face validity of the measure and to align the measure with the older adult and pediatric PQI/PDI for asthma/COPD admissions. Patients are identified without additional burden.
- Patients with severe chronic respiratory diseases have been excluded.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

- Statistical risk model with 23 risk factors

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

The developer reports:

- The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect)
  - The covariates are gender and age (in 5-year age groups)
  - An option model is available that includes percent of households under the federal poverty level
  - A conceptual model acknowledging the impact of community factors also was considered (e.g., clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise, community norms and beliefs, etc., which can impact hospitalization rates).
- The risk adjustment model was calibrated by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.
  - The developer reports the observed to predicated values across the deciles range between 0.88-1.12, indicating it is well calibrated.
  - The developer reports the c-statistic (measure of well the risk adjustment model distinguishes events from non-events, is the c-statistic) was 0.56, which it concludes was poor and presumes was due to the limited predictors included.
  - The developer noted the addition of SES to the model improves the calibration (range of 0.95-1.03), but that the c-statistic did not change substantially.

Questions for the Committee:

- *Is the risk adjustment methodology appropriate?*
- *Were the appropriate community factors included in the conceptual model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance*

measure scores can be identified):

The developer assessed the probability that a county is higher or lower than a benchmark or threshold, given county size—i.e., whether the indicator can discriminate the best performing counties from the lower performing counties. The developer reports on discrimination results for counties:

- Measure has poor discrimination to:
  - identify low performing hospitals for most hospitals: 38% of counties can be classified as better or worse than the threshold
  - identify high performing hospitals; 33% of counties can be classified as better or worse than the benchmark.
- Performance discrimination remains strong when adding SES to risk adjustment.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

The developer notes:

- The AHRQ QIs use frequently reported administrative data variables. PQI 11 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis.
- Missing data handled – rates are typically less than 1% of the state database.

Guidance from the Validity Algorithm : 1 → 2 → 3 → 6 → 7 → 8 (highest eligible rating is HIGH)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**specifications are consistent with the evidence

**I find no discrepancies that might be labeled inconsistant.

**The specifications are consistent with the target populations and the results should be meaningful. Need to continue to monitor though as the improvements demonstrated in the data were small.

**The validity situation is not so strong. There is not prior validity data to compare to. There is a type (COPD versus asthma) in the description of validity testing which undermines confidence in this section. There is a similar typo earlier in the document where the word pneumonia appears to have been substituted for asthma. While there is some reason to believe that the measure is meaningful, including the progressive improvement over time in the measure that possibly reflects improvements in care and provider behavior, the statistical measure provided is significant for some factors (prevalence, behaviors, SES and environment) but not for certain others.

**Yes

**Face validity and empirical validity tested and appreciate the documentation of correlation between

2a2. Reliability Testing

Comments:

**face validity measured with input from four clinical expert panels, panel conclusions reasonable. empirical validity testing indicated prevalence, health behaviors, and SES/environment all SS predictors ($p<0.0001$) although unclear why copd (not asthma) used fro empirical validity assessment

**Yes.

The level of validity testing conducted included the following:

Empirical validity testing

Systematic Assessment of Face Validity

The clinical panel noted that this indicator overall was useful. They identified specific actions that could improve rates, especially from a population health perspective, such as access to medications, patient education, reduction of risk factors, such as environmental exposure to pollution or allergens and smoking. However, it is important, as with many population health indicators to acknowledge the complex factors influencing the indicators.

SES explained the most variance, however moderate correlation between socioeconomic status and other factors suggest that it is

difficult to fully ascertain the unique contribution of each factor on asthma hospitalizations. The disparities table (see Table 2 in the supplemental files) demonstrates that zip codes in the highest income quartile have 31% lower admission rates than those in the lowest income quartile.
From a population health perspective such disparities argue for the importance of the indicator in capturing poor outcomes for vulnerable populations. Further, taking a population health perspective, efforts to decrease individual risk factors such as obesity, smoking and limited physical exercise (the variables included in the health behaviors factor) may decrease prevalence and exacerbations of asthma.
**The measure demonstrated validity.
**Sufficient confidence in validity exists to justify widespread implementation with the caveat that it will not be reliable for communities with less than 3800 members. It is not purely a measure of quality. It reflects various environmental factors and background or structural issues that will not be impacted by improvements in the quality of care.
**Moderate validity. Concerned by statement that "access to care was not significant when HB and SES are included in the model"
**Validity tested both face and empirically and results rank high which would allow for conclusions to be made re: quality of care

2b2. Validity Testing
Comments:
**exclusions c/w evidence
risk adjustment model well calibrated, but c-statistic is poor suggesting additional predictors should be included.
regarding missing data as a threat: never exceeds 1% of eligible records, no inappropriate exclusions, all exclusions c/w the evidence.
**Using the 2013 data from 40 states, the developers examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.
A total of 262 discharges were excluded due to diagnoses of cystic fibrosis and anomalies of the respiratory system. Removing this exclusion would increase the numerator count by 0.13%. The denominator does not change. Although discharges transferred into a hospital are excluded, these encounters are captured in the area level numerator via the originating hospitalization.
The exclusion of cystic fibrosis and anomalies of the respiratory system has been retained to increase the face validity of the measure and to align the measure with the older adult and pediatric PQI/PDI for asthma/COPD admissions, although the impact on the numerator is minor for PQI15. Patients meeting the exclusion are identifiable without additional burden using the same data as is used to identify numerator qualifying discharges.
The indicator excludes patients with severe chronic respiratory diseases, because asthma in the context of these diseases differs clinically from the patients with asthma alone.

_____
2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.
2b4.1. What method of controlling for differences in case mix is used?
☐ No risk adjustment or stratification
☒ Statistical risk model with 23 risk factors
☐ Stratification by Click here to enter number of categories risk categories
☐ Other, Click here to enter description
2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not applicable
2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factor without additional burden using the same data as is used to identify numerator qualifying discharges.
The indicator excludes patients with severe chronic respiratory diseases, because asthma in the context of these diseases differs clinically from the patients with asthma alone.
Exclusion of cases for missing data is appropriate.
**Exclusions are appropriate for the measure.
**no
**2b3. Yes, exclusions consistent with evidence. Do wonder about whether COPD should be excluded.
No inappropriate exclusions.
2b4. Risk assessment seems appropriate. Appropriate community factors are included in the model. Not clear then if the proposal is to risk adjust this measure. Is there a signal for quality if risk-adjusted?
2b5. Per report, poor discrimination to identify meaningful differences about quality.
2b6. NA
2b7. No issue.
**Missing data handled consistently throughout the AHRQ PQI measures being evaluated for endorsement maintenance by this

workgroup

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**reliability testing demonstrated good reliability with signal-to-noise of 0.75 when county size included > 3800 eligible members, when SES added to risk adjustment, reliability remains adequate at 0.74
**Yes.
The indicator demonstrates good reliability with a signal-to-noise ratio of 0.75. Reliability meet threshold for all counties, except for the smallest counties with eligible populations under approximately 3800 individuals. Smoothed rates, which are recommended for all counties (and are implemented in the AHRQ software), address any reliability concerns for the smallest counties. When SES is added to the risk adjustment, the reliability remains adequate at 0.74.
**Reliability in the measure was tested and proven. Results should be consistent at the data level.
**Reliability was tested at the score level only.
**Reliability: 4. Moderate
**Signal to noise testing using entire US population data for calculated rates, and achieved a high score for reliability. No questions about methods or results and believe this would rate as high

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer reports:

- All data elements are in defined fields in electronic claims
- The measure is based on readily available administrative billing and claims data.
- There are no fees. The AHRQ QI software is publically available from the AHRQ Quality Indicators website.
- Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**yes, required data readily available and can be implemented for performance measurement, no fees, software freely available
**All data elements are generated routinely and all are publically available in electronic form as no cost.
**Administrative data so it is a feasible measure. ICD-10 may create some challenges in updating the measure and ensuring consistent results.
**This is electronically implemented and is thus quite feasible.
**High
**No concerns, all see feasible

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

The measure is currently being used in the following programs:

- Public Reporting:
  - Arizona Department of Health Services, AZ Hospital Compare
  - Connecticut Department of Health Services, CT Hospital Compare
  - Maine Health Data Organization (MHDO),
  - Nevada Compare Care
  - Oklahoma State Department of Health, MONAHRQ
  - Utah Department of Health, MONAHRQ website
  - Virginia Health Information, MONAHRQ website
  - Washington State, MONAHRQ website
  - California Office of Statewide Health Planning and Development, Healthcare Information Division
  - Connecticut, Office of Health Care Access
  - Arkansas Department of Human Services: Arkansas Medicaid Performance
  - Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
  - CMS Medicaid Adult Core Measures
- Payment programs:
  - CMS Medicare Shared Savings Program
  - Oregon Health Authority
- Regulatory and Accreditation Programs
  - Statewide Quality Advisory Committee (Massachusetts): Center for Health Information and Analysis
- Quality Improvement with benchmarking:
  - West Jefferson Medical Center

Improvement results

- The PQI 11 hospital admissions rate has decreased by 9,000 fewer hospitalizations from 2011-2013
  - In 2012 the rate was 0.5 per 1,000
  - in 2013 the rate was 0.4 per 1,000
- Variation among counties decreased substantially in 2013. The developer is not certain whether this is a single year anomaly or an actual trend.

Unexpected findings (positive or negative) during implementation

- No challenges implementing this measure. This use of the AHRQ PQIs has grown since the initial endorsement, suggesting the measure is implementable and useful.
- The transition to ICD-10-CM will provide challenges in understanding time trends or rates from calendar year 2015. This is a challenge all measures based on coded data will encounter.

Potential harms: The developer did not identify potential harms.

Feedback : No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Question for the Committee**:*

- *Given the relatively small improvement from 2011-2013. can the performance results be used to further the goal of high-quality, efficient healthcare?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement
4c. Unintended Consequences
Comments:
**measure currently publicly reported, currently in use in several
accountability program
performance results can be used to improve hospital admission rates, and decrease variability among counties.
no unexpected findings, challenges associated with change from ICD-9 to ICD-10 as with all measures relying on ICD coded data
**The measure is publicly reported by a wide randge of health services and paymen providers.
**It is being used quite broadly, including being reported by 13 or more health departments. There do not appear to be any issues with implementation or unintended consequences. There have been some improvements documented by the measure. The sensitivity of the measure to changes in quality may be diluted because of the background issues that are not modified by changes in quality of care.
**Moderate. The small improvement from 2011 to 2013 despite its widespread adoption does question the idea that the results can be used to further the goal of high quality, efficient healthcare.
**Useful and currently being used, calculated and reported through AHRQ

Criterion 5: Related and Competing Measures
Related or competing measures • 0728: Asthma Admission Rate (PDI 14)

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0283

Measure Title: Asthma in Younger Adults Admission Rate (PQI 15)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF***

staff if more pages are needed.

- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Hospitalization for asthma

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Asthma among young adults age 18-39, for the most part, can be treated in the outpatient setting. Numerous studies have shown an association at the patient level, between appropriate treatment and hospital admission rates, although education interventions have not always had an impact on admission rates.[1,2] Asthma admission rates are also associated conceptually, and in some cases empirically, with community-level pollution[3] and smoking rates.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

This indicator is intended to identify hospitalizations for asthma in younger adults age 18-39. With appropriate pharmaceutical and other outpatient management, risk of hospitalization is decreased.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Please note that this is an outcome measure. Therefore, a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.4.1 and 1a.8 below, to provide additional context and support for the measure.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not applicable

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches are performed to continually update evidence for all AHRQ PQIs, including PQI15. The current evidence review results presented below constitute articles published from January 2012 - October 2015. Additional articles from earlier years were identified from previous project-related literature reviews. Search terms included the relevant MeSH term (asthma). We combined this clinical search string with hospital*[Title/Abstract]) AND (prevent*[Title/Abstract] OR "access to care"[Title/Abstract] OR "ambulatory care sensitive"[Title/Abstract] OR “avoidable hospitalization”[Title/Abstract] OR "small area analysis"[MeSH]). For completeness, we also tested more inclusive search strings. Below, we have provided a summary of the most up-to-date evidence.

1a.8.2. Provide the citation and summary for each piece of evidence.

This section describes the current evidence related to PQI 15. Because studies evaluating asthma exacerbations often combine Emergency Department (ED) visits with hospitalizations, we note below when this combined outcome was assessed. Furthermore, asthma studies include patients of varying ages, and may combine data from children and young adults or young adults with older adults. We include in the evidence for PQI 15, studies with various age ranges that include young adults.

Importance

Using 2000-2010 NIS and SID data, one HCUP statistical brief by the Agency for Health Care Policy and Research (2014) examined geographic and temporal variation in PQI15 (version 4.1) rates.[4] The rate of adult asthma-related hospital stays remained relatively unchanged between 2000 and 2010, at about 119 hospital stays per 100,000 population. Adults treated in the Northeast had a higher rate of hospital stays for asthma (167.6 per 100,000 population) than adults in other Census regions (which were 110.0 or less per 100,000 population) ($p < 0.05$). A study by Roy et al. (2010) found that asthma hospitalization rates were significantly higher among all demographic groups in the rural Delta region compared with the urban Jackson Metropolitan Statistical Area ($P < 0.001$). Residents of the Delta also had higher odds for multiple hospitalizations when controlling for race, sex, age, and household income ($P < 0.05$).[5] In another study of an urban Philadelphia population, rates of hospitalization declined significantly between 1995–1997 and 1997-1999 (from 23 to 19 per 10,000; $\chi^2 = 30.62$; $P < .001$).[6]

Validity

As a population health indicator, we consider the relationship of PQI 15 to a wide range of factors that are amenable to changes in public policy and community based interventions that may in turn improve access to quality care and community resources, reduce risky personal behaviors or improve self-care, reduce environmental exposure, or prevent the development of asthma.

Disparities

Disparities in asthma hospitalization rates among non-whites, particularly among Blacks, has been documented in numerous studies. Using 2000-2010 NIS and SID data, the Agency for Health Care Policy and Research revealed gender, racial and income-based disparities in PQI15 (version 4.1) rates. [4] In 2010, females had a 129 percent higher rate of hospital stays than males (163.0 versus 71.2 hospital stays per 100,000 population), while African American and Hispanic patients had higher rates of asthma hospitalization (297.9 and 144.6 per 100,000 population, respectively) than White and Asian and Pacific Islander patients (90.5 and 65.4 per 100,000 population, respectively). Additionally, adult patients in the lowest income communities had higher rates of hospital stays for asthma than those in the highest income communities (194.3 versus 72.6 hospital stays per 100,000 population). All differences noted exhibited at least a 10 percent difference between estimates and were statistically significant at 0.05 or better. [4] In a large prospective cohort study, Blacks and Asians had higher asthma hospitalization rates compared to Whites (relative risks (RR) 1.7; 95% CI 1.4–2.0 and RR 1.6; 95% CI 1.2– 2.1), but Hispanics did not (RR 0.9; 95% CI 0.6–1.4). Among Asians, increased risk was concentrated in Filipino men and women and South Asian men.[8] In another study of an urban population, African Americans had signifigicantly higher hospitalization rates than Caucasians (45.7 vs. 7.6 per 10,000). These findings were consistent across all poverty levels. Moreover, asthma hospitalization was significantly associated with poverty area residence (relative risk [RR], 2.29) and with African American race (RR 4.31) (both $p < 0.001$).[6] Roy et al. (2010) found that hospitalization rates were higher among blacks and females ($p < 0.001$). In both Mississipi regions studied, Blacks were more likely than Causasians to have 3 or more asthma hospitalizations ($P < .001$).[5] In a retrospective analysis, asthma hospitalizations were associated with male gender (OR 0.67; 95% CI 0.52–0.86) and with residence in neighborhood in which more than 10% of the population is non-white (OR 1.62; 95% CI 1.23–2.13).[9] Sawicki et al. (2010) reported that asthma hospitalizations were associated with residence in neighborhood in which more than 50% of adults have only high school education or less (OR 1.39; 95% CI1.02–1.89).[9] By contrast, another study by Gold et al. (2013) found that across all levels of asthma control, non-whites did not have significantly higher rates of hospitalizations than whites.[7]

Community characteristics

The following section discusses evidence related to leverage points within the community outside the healthcare system. It describes the relationship of county characteristics to asthma hospitalization. Interventions which effectively modify these characteristics, may impact asthma hospitalization. Although, all studies were correlational in nature, and cannot speak to the causal nature of the relationship.

Environmental exposure

Silverman and Ito (2010) found an increased risk for total asthma hospitalizations associated with both Particulate Matter ($PM_{2.5}$) and ozone. These estimated risks were age-dependent, with the stronger associations appearing for those under age 19 than for those over the age of 18. The risk of non-ICU asthma hospitalization among all age groups for $PM_{2.5}$ and ozone were each 1.09 (95% CI 1.06-1.12).[10] Anderson et al. (2012) observed the effect of traffic-related air pollution on older Danish adults over time and found that NO_2 levels were associated with risk for asthma hospitalization (HR and 95% CI per IQR, 5.8 $\mu g/m^3$: 1.12; 1.04-1.22), and for first-ever admissions (1.10; 1.01-1.20), with the highest risk in people with a previous asthma hospitalization (1.41; 1.15-2.07) ($p < 0.05$, Wald test for interaction).[11] Another small study observed 142 workers exposed to welding and found that welding exposure was the fifth leading cause of work-related asthma, a condition which required hospitalization in 36.7% (n=50) of the study sample.[12]

Access to care

The following section discusses evidence related to leverage points within the health care system. The following studies primarily examine alternative care models and patient-level hospitalization rates. While it cannot be assumed from these studies that improving care for patients will necessarily result in lower area-level hospitalization rates, these studies identify potential mechanisms to improve outcomes for patients.

Several studies have linked asthma hospitalizations with other markers of asthma severity or uncontrolled disease. Gold et al. (2013) examined various patient characteristics and reported that those with asthma characterized as uncontrolled or partly controlled had more hospitalizations than patients whose asthma was well-controlled (mean 12-month hospitalizations were 0.5, 0.07 and 0.03, respectively; $p < 0.001$).[7] In one study of an urban population, asthma hospitalizations were directly correlated with prescriptions for inhaled short-acting β-agonists in 1995–1997 and 1997-1999 (r_s = 0.61 and 0.60 respectively for both time periods, $p < 0.001$) and inversely correlated with long-acting β-agonists (LABA) prescriptions during the same time periods (r_s = -0.56 and -0.66, respectively, $p < 0.001$). This study also found higher hospitalization rates for 35- to 64-year-olds than for 18- to 34 year olds (13.7 vs. 11.9 per 10,000); and asthma hospitalization was also significantly associated with age (RR 1.15; P <.001).[6]

However, Williams et al. (2011) found that an estimated 24% of asthma exacerbations were attributable to inhaled corticosteroid (ICS) medication nonadherence. In that study, inhaled corticosteroid adherence varied in the time period leading up to an asthma exacerbation and was associated with a reduction in asthma exacerbations (including hospitalization), but this association was only statistically significant among patients whose adherence was greater than 75% of the prescribed dose (Hazard ratio 0.61; 95% CI 0.41-0.90) when compared with patients whose adherence was 25% or less. This pattern was largely confined to patients whose asthma was not well controlled initially.[13]

Schlender et al. (2012) developed a model that predicted the clinical effects of two widely used controller medications, ICS and LABA, and then applied that model to a population derived from the National Asthma Survey to quantify the effects of increasing prescriptions to guideline-recommended levels, increasing adherence, or both, on the frequency of hospitalizations.[14] They used the simulation model to estimate the impact of increased corticosteroid use under EO (Expanded Prescribing Observed Adherence), OP (Observed Prescribing Perfect Adherence), and EP (Expanded Prescribing Perfect Adherence) scenarios on averting hospitalizations. The authors found that differences in outcomes rates observed for EO and OP scenarios were not significant, however, the mean hospitalizations, per year for the EP group was 0.04 (0.01 SD) compared to the OO (Observed Prescribing Observed Adherence) group with a mean of 0.17 (0.03 SD); $p < 0.03$. Under the EP scenario, the model predicts a 1,100,000 (80%) of overnight hospitalizations would be averted relative to OO.

Several studies examine the efficacy of interventions aimed at reducing rates of adult asthma hospitalizations. A Cochrane review of 5 studies demonstrated that educational interventions for adults who visit the emergency room for acute asthma led to a reduction in the risk of subsequent hospitalizations (RR 0.50; 95% CI 0.27-0.91; N=572).[2] Chamnan et al. (2010), in examining a small study set in Thailand, found that among the 57 patients enrolled in a 12-week disease management program, hospitalizations with acute asthma attacks decreased from 0.14 to 0.04 per patient ($p = 0.034$) following the intervention.[15] However, three studies found limited impact for asthma education efforts. One study examining the effects of home-based education programs, found no significant impact of interventions on asthma

hospitalizations among Medicaid-managed care patients.[1] Mancuso et al (2010) tested the effect of an educational intervention designed to improve asthma knowledge and self-efficacy among patients with depressive symptoms in a primary care setting, and found no difference in asthma hospitalizations between treatment and control groups.[16] Apter et al. (2011) investigated whether an individualized problem-solving intervention improves asthma outcomes over adult education efforts, and found no significant differences between the two groups in hospitalization rates.[17]

Seasonal variation, influenza and vaccination

Asthma-related healthcare utilization is also affected by seasonal factors. One study by Gerke et al. (2014) used NIS data (1998-2008) to reveal that asthma hospitalization rates with a secondary diagnosis of influenza were significantly associated with elevated influenza activity ($p<0.0001$).[18] Trogdon et al. (2010) used MEPS data to show that adults with asthma vaccinated for influenza were 4.4 percentage points less likely to have an inpatient stay due to acute and chronic respiratory conditions (95% CI = -10.8 to -1.0).[19]

Fitzgerald et al. (2014) examined data from New York State to track seasonal variation of adult asthma-related hospital admissions.[20] The authors found asthma admissions in the state decreased during cold spells in the winter months of December through March (-4.9% decline in mean daily asthma admissions, 95% CI -7.83 to -1.88). This decline was stronger in upstate New York (-5.51% change, 95% CI -9.52 to -1.33), which is generally colder than the rest of the state. A different pattern was evident after a cold spell for the transitional months of November and April. In both of these months, there was a significant increase in asthma hospitalizations after a cold spell in New York State (9.63% increase in November, 95% CI 5.51 to 13.92 and a 5.00% increase in April, 95% CI 1.19 to 8.96) and in upstate New York specifically (8.46% increase in November, 95% CI 0.58 to 16.95 and a 7.70% increase in April, 95% CI 0.08 to 15.89).

References

1. Catov JM, Marsh GM, Youk AO, Huffman VY. Asthma home teaching: two evaluation approaches. Dis Manag. Jun 2005;8(3):178-187.
2. Tapp S, Lasserson TJ, Rowe B. Education interventions for adults who attend the emergency room for acute asthma. The Cochrane database of systematic reviews. 2007(3):CD003000.
3. Sheppard L, Levy D, Norris G, Larson TV, Koenig JQ. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994 Epidemiology. 1999;10(1):23-30.
4. Barrett ML, Wier LM, Washington R. Trends in Pediatric and Adult Hospital Stays for Asthma, 2000-2010: Statistical Brief #169. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Health Care Policy and Research (US); 2014.
5. Roy SR, McGinty EE, Hayes SC, Zhang L. Regional and racial disparities in asthma hospitalizations in Mississippi. Journal of Allergy and Clinical Immunology. 2010;125(3):636-642.
6. Lang DM, Polansky M, Sherman MS. Hospitalizations for asthma in an urban population: 1995–1999. Annals of Allergy, Asthma & Immunology. 2009;103(2):128-133.
7. Gold LS, Yeung K, Smith N, Allen-Ramey FC, Nathan RA, Sullivan SD. Asthma control, cost and race: results from a national survey. The Journal of asthma : official journal of the Association for the Care of Asthma. Sep 2013;50(7):783-790.
8. Tran HN, Siu S, Iribarren C, Udaltsova N, Klatsky AL. Ethnicity and risk of hospitalization for asthma and chronic obstructive pulmonary disease. Annals of epidemiology. 2011;21(8):615-622.
9. Sawicki GS, Vilk Y, Schatz M, Kleinman K, Abrams A, Madden J. Uncontrolled asthma in a commercially insured population from 2002 to 2007: trends, predictors, and costs. Journal of Asthma. 2010;47(5):574-580.
10. Silverman RA, Ito K. Age-related association of fine particles and ozone with severe acute asthma in New York City. Journal of Allergy and Clinical Immunology. 2010;125(2):367-373. e365.

11. Andersen ZJ, Bønnelykke K, Hvidberg M, et al. Long-term exposure to air pollution and asthma hospitalisations in older adults: a cohort study. Thorax. 2012;67(1):6-11.

12. Banga A, Reilly MJ, Rosenman KD. A study of characteristics of Michigan workers with work-related asthma exposed to welding. Journal of Occupational and Environmental Medicine. 2011;53(4):415-419.

13. Williams LK, Peterson EL, Wells K, et al. Quantifying the proportion of severe asthma exacerbations attributable to inhaled corticosteroid nonadherence. Journal of Allergy and Clinical Immunology. 2011;128(6):1185-1191. e1182.

14. Schlender A, Alperin PE, Grossman HL, Sutherland ER. Modeling the Impact of Increased Adherence to Asthma Therapy. PloS one. 2012;7(12):e51139.

15. Chamnan P, Boonlert K, Pasi W, et al. Implementation of a 12-week disease management program improved clinical outcomes and quality of life in adults with asthma in a rural district hospital: pre-and post-intervention study. Asian Pacific Journal of Allergy and Immunology. 2010;28(1):15-21.

16. Mancuso CA, Sayles W, Allegrante JP. Randomized trial of self-management education in asthmatic patients and effects of depressive symptoms. Annals of Allergy, Asthma & Immunology. 2010;105(1):12-19.

17. Apter AJ, Wang X, Bogen DK, et al. Problem solving to improve adherence and asthma outcomes in urban adults with moderate or severe asthma: a randomized controlled trial. Journal of Allergy and Clinical Immunology. 2011;128(3):516-523. e515.

18. Gerke AK, Yang M, Tang F, Foster ED, Cavanaugh JE, Polgreen PM. Association of hospitalizations for asthma with seasonal and pandemic influenza. Respirology (Carlton, Vic.). Jan 2014;19(1):116-121.

19. Trogdon JG, Nurmagambetov TA, Thompson HF. The economic implications of influenza vaccination for adults with asthma. American journal of preventive medicine. 2010;39(5):403-410.

20. Fitzgerald EF, Pantea C, Lin S. Cold spells and the risk of hospitalization for asthma: New York, USA 1991-2006. Lung. Dec 2014;192(6):947-954.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PQI15_NQF_0283_Measure_Evidence_Form_151214.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This indicator is intended to identify hospitalizations for asthma in younger adults age 18-39. With appropriate pharmaceutical and other outpatient management, risk of hospitalization is decreased.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included).*

This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.
This table is also included in the supplemental files.
Table 1. Reference Population Rate and Distribution of County Performance for PQI 15

Overall Reference Population Rate

Year	Number of Counties	Number of Events (Numerator)a	Population at Risk (Denominator)a	Observed Rate Per 1,000a
2009	3,135	52,986	91,475,217	0.5792
2010	3,138	46,476	91,767,953	0.5065
2011	3,141	43,685	92,184,336	0.4739
2012	3,139	43,746	90,798,464	0.4818
2013	3,140	34,549	91,667,214	0.3769

Distribution of County-level Observed Rates in Reference Population Per 1,000

Year	Number of Counties(p=percentile)b	Mean	SD	p5	p25	Median	p75	p95
2009	3,135	0.50	0.63	0.00	0.00	0.38	0.70	1.51
2010	3,138	0.46	0.63	0.00	0.00	0.34	0.64	1.38
2011	3,141	0.40	0.53	0.00	0.00	0.29	0.56	1.25
2012	3,139	0.67	17.85	0.00	0.00	0.26	0.51	1.07
2013	3,140	0.28	0.37	0.00	0.00	0.18	0.42	0.90

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)
aThe observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible (ages 18-39) population of all counties included in the reference population data (denominator). Note: Observations from counties with rates outside of 1.5*interquartile range are excluded as outliers.
bThe distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile. Note: Counties with rates outside of 1.5*interquartile range are excluded as outliers.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
n/a

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental files.

Table 2. Admission Rates per 1,000 (PQI 15), by patient and hospital characteristics, 2013

Patient/hospital characteristic	Estimate	Std Error	p-value	(Ref Grp = *)	Lower 95% CL	Upper 95% CL
Total U.S.	37.69	0.2027			37.29	38.09
Patient Characteristics						
Age Groups:						
18-39	37.70	0.2028			37.30	38.10
40-64						
65 and over						
Gender:						
Male*	25.11	0.2855			24.55	25.67
Female	50.47	0.2879	<.001		49.91	51.04
Patient Zip Code Median Income						
First quartile (lowest income)	48.05	0.7265	<.001	46.63	49.48	

Second quartile	45.80	0.5079	<.001	44.80	46.79		
Third quartile	38.66	0.4207	<.001	37.84	39.49		
Fourth quartile (highest income)*	33.31	0.2783		32.76	33.85		
Location of patient residence (NCHS):							
Rural	16.88	1.4584	<.001	14.02	19.74		
Urban*	38.10	0.2047		37.70	38.50		
Location of Care:							
Northeast*	46.318	0.484		45.37	47.27		
Midwest	46.835	0.441	0.215	45.97	47.70		
South	34.867	0.333	<.001	34.21	35.52		
West	28.511	0.412	<.001	27.70	29.32		

Source: Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project, Nationwide Inpatient Sample, 2013, and AHRQ Quality Indicators, version 6.0.
Rates are adjusted by age and gender using the AHRQ QI PQI Reference Population for 2013 as the standard population; when reporting is by age, the adjustment is by gender only; when reporting is by gender, the adjustment is by age only.
NCHS - National Center for Health Statistics designation for urban-rural locations.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
n/a

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Asthma is one of the most common chronic lower respiratory diseases. In 2012, an estimated 18.7 million adults had asthma.(1) Asthma results in 1.8 million emergency department visits in 2011(2) and was responsible for 3,311 deaths in 2012 (age-adjusted rate of 1.0 per standard population of 100,000).(3) While hospital stays due to asthma declined in the pediatric population between 2000 and 2010, asthma-related hospital stays for all adults remained stable at 119 per 100,000 population. The average cost of each hospitalization increased from $5,200 to $6,600.(4)

Table 1 shows that in the AHRQ QI 2013 PQI Reference Population there were 34,549 qualifying discharges with a rate of 0.4 per 1,000. The coefficient of variation was 1.32 in 2013.(5)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. National Center for Health Statistics. Vital Health Stat 10(260). 2014.
2. National Hospital Ambulatory Medical Care Survey: 2011 Emergency Department Summary Tables. Available at http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf. Last accessed December 07, 2015
3. Hoyert DL, Xu JQ. Deaths: preliminary data for 2011. Natl Vital Stat Rep. 2012;61(6):1-65. Hyattsville, MD: National Center for Health Statistics.2012.
4. Barrett ML, Wier LM, Washington R. Trends in Pediatric and Adult Hospital Stays for Asthma, 2000-2010. HCUP Statistical Brief #169. January 2014. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb169-Asthma-Trends-Hospital-Stays.jsp. Accessed December 4, 2015.
5. HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
n/a

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Prevention, Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Access, Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://1.usa.gov/1JpQ10E Note: The URL link currently provides Version 5.0 specifications. Version 6.0 specifications will be released publicly March 2016 and found via the module page: http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PQI15_Technical_Specifications_v6.0_151214_v02.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
As standard protocol, the AHRQ QI program annually updates all measures with Fiscal Year coding changes, refinements based on stakeholder input, refinements to improve specificity and sensitivity based on additional analyses, and necessary software changes. In addition, approximately every two years, AHRQ updates the risk adjustment parameter estimates and composite weights based on the most recent year of data (i.e., the most current reference population possible). The refined measures are tested and confirmed to be valid and reliable prior to release of the updated software.

Since the last update, the following changes have been made to the indicator:
- Added exclusion for cases with any listed ICD-9-CM diagnosis codes for cystic fibrosis and anomalies of the respiratory system.
- Changed eligible age range for numerator and denominator to age 18-39 following clinical panel review. Asthma in older adults (ages 40 and above) is captured in PQI 05.
- Removed exclusion for discharges in MDC 14 and MDC 15. By definition, discharges with a principal diagnosis of asthma are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QITM software does not explicitly exclude obstetric cases.
- The data upon which to base the reference population was updated.
- Updated with 2013 US Census population estimates
- Fiscal Year coding updates (none in this time period)

Additional information regarding revisions to PQI software and technical specifications available online:
http://www.qualityindicators.ahrq.gov/Modules/pqi_resources.aspx
Note: Version 6.0 specifications will be released publicly March 2016.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Discharges, for patients ages 18 through 39 years, with a principal ICD-9-CM or ICD-10-CM/PCS diagnosis code for asthma.

[NOTE: By definition, discharges with a principal diagnosis of asthma are precluded from an assignment of MDC 14 by grouper software. Thus, obstetric discharges should not be considered in the PQI rate, though the AHRQ QI software does not explicitly exclude obstetric cases.]

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Users may specify a time period; but the time period is generally one year. Note that the reference population rates and signal variance parameters assume a one-year time period.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Please see attached excel file in S.2b. for Version 6.0 specifications.

Prevention Quality Indicators technical specifications and appendices also available online at http://www.qualityindicators.ahrq.gov/Modules/PQI_TechSpec.aspx). Note: The URL link currently provides Version 5.0 specifications. Version 6.0 specifications will be released publicly March 2016.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Population ages 18 through 39 years in metropolitan area or county. Discharges in the numerator are assigned to the denominator based on the metropolitan area or county of the patient residence, not the metropolitan area or county of the hospital where the discharge occurred.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
† The term "metropolitan area" (MA) was adopted by the U.S. Census in 1990 and referred collectively to metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs) and primary metropolitan statistical areas (PMSAs). In addition, "area" could refer to either 1) FIPS county, 2) modified FIPS county, 3) 1999 OMB Metropolitan Statistical Area or 4) 2003 OMB Metropolitan Statistical Area. Micropolitan Statistical Areas are not used in the QI software.

See AHRQ QI website for 2014 Population File Denominator report for calculation of population estimates embedded within AHRQ QI software programs. http://www.qualityindicators.ahrq.gov/Downloads/Software/SAS/V50/AHRQ_QI_Population_File_V50.pdf

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Not applicable.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Not applicable.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups). An option model is available that includes percent of households under the federal poverty level as well. Because we cannot individually observe the age and gender of each person in a counties population, we use the age and gender distribution of the county to estimate the number of "cases" in each age*gender group. The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Data (SID) for the year 2013 (combined), a database consisting of 40 states and the U.S. Census data by county. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., area). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Additional information on methodology can be found in the Empirical Methods document on the AHRQ Quality Indicator website (www.qualityindicators.ahrq.gov) and in the attached supplemental information.

The specific covariates for this measure are as follows:

PARAMETER	LABEL
SEX	Female
AGE	Male, Age 18-24
AGE	Male, Age 25-29
AGE	Male, Age 30-34
AGE	Male, Age 35-39
AGE	Female, Age 18-24
AGE	Female, Age 25-29
AGE	Female, Age 30-34
AGE	Female, Age 35-39
POVCAT	Poverty Decile 2
POVCAT	Poverty Decile 3
POVCAT	Poverty Decile 4
POVCAT	Poverty Decile 5
POVCAT	Poverty Decile 6
POVCAT	Poverty Decile 7
POVCAT	Poverty Decile 8
POVCAT	Poverty Decile 9
POVCAT	Poverty Decile 10 (Highest percent poverty)1

1Deciles are based on the percentage of households under the federal poverty level (FPL).
Source: http://qualityindicators.ahrq.gov/Modules/pqi_resources.aspx
Parameter estimates with and without SES covariates (POVCAT) are included with the Technical Specifications.
Please note Version 6.0 will be released publicly March 2016.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 observed rate of each PQI is simply the number of individuals living in a county admitted to the hospital for the condition of interest divided by the census population estimate for the area (for PQI 15 ages 18-39). The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected performance observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic distributions observed in the user's dataset? The expected rate is calculated only for risk-adjusted indicators.

The expected rate is estimated for each county using logistic regression.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the performance observed in the user's dataset were applied to a mix of patients with demographics distributed like the reference population. The risk adjusted rate is calculated using the indirect method as observed rate divided by expected rate multiplied by the reference population rate. The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural counties).

For additional information, please see supporting information in the Quality Indicator Empirical Methods attached in the supplemental files.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing), or county (PSTCO=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications and software are specified to be used with any ICD-9-CM or ICD-10-CM/PCS coded administrative billing/claims/discharge dataset.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Population : County or City

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: All community based care

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
n/a

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PQI15_NQF_0283_Measure_Testing_Form_151214v02.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0283
Measure Title: Asthma in Younger Adults Admission Rate (PQI 15)
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.

- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2009-2013. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ)[1]. HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 40 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

The SID data elements include ICD-9-CM coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

For additional testing of the indicators we created a county characteristic dataset which was merged with the county-level observed and risk adjusted rates from the HCUP dataset described above. Using a conceptual model for hospitalization indicators, we examined candidate predictor variables from the County Health Rankings (CHR)[2] dataset and the American Community Survey (ACS)[3]. Candidate variables that corresponded to the conceptual model were grouped by the following categories: individual health behavior (IHB) included variables that include actions and behaviors of individuals and may be mutable, access to care (AC) included variables that reflect the structure and quality of the healthcare system in a community, socioeconomic status (SES) included variables of poverty and education levels within a community and environment (E) variables included community characteristics, such as access to food and open spaces, pollution or violent crime. County prevalence estimates were derived from the Behavioral Risk Factor Surveillance System (BRFSS) Asthma model based county estimates.

1.3. What are the dates of the data used in testing?

HCUP data: 2009-2013, CHR data: 2014 and ACS data: 2013.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

---

[1] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

[2] "County Health Rankings & Roadmaps". University of Wisconsin Population Health Institute, 2014. http://www.countyhealthrankings.org. Accessed 26 Jan. 2015.

[3] "American Community Survey (ACS)." *https://www.census.gov/programs-surveys/acs/data.html*. Accessed 2015.

County

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Population Health	☒ other: County

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1. Reference Population Rate and Distribution of County Performance for PQI 15 Asthma in Younger Adults Admission Rate

Overall Reference Population Rate				
Year	Number Counties	Outcome of Interest (Numerator)[1]	Population at Risk (Denominator)[1]	Observed Rate Per 1000[1]
2009	3,135	52,986	91,475,217	0.5792
2010	3,138	46,476	91,767,953	0.5065
2011	3,141	43,685	92,184,336	0.4739
2012	3,139	43,746	90,798,464	0.4818
2013	3,140	34,549	91,667,214	0.3769

Distribution of County-level Observed Rates in Reference Population Per 1000								
Year	Number of Counties	(p=percentile)[2]						
		Mean	SD	p5	p25	Median	p75	p95
2009	3,135	0.50	0.63	0.00	0.00	0.38	0.70	1.51
2010	3,138	0.46	0.63	0.00	0.00	0.34	0.64	1.38
2011	3,141	0.40	0.53	0.00	0.00	0.29	0.56	1.25
2012	3,139	0.67	17.85	0.00	0.00	0.26	0.51	1.07
2013	3,140	0.28	0.37	0.00	0.00	0.18	0.42	0.90

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population for the county in the reference population data (denominator). For PQI 15 this includes population ages 18-39 years.

[2]The distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

See 1.5 (Table 1)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The reliability and performance discrimination testing was completed using 2013 HCUP data. Validity testing was completed using 2012 HCUP data. The AHRQ QI PQI 2012 reference population has 34,440,38 discharges and the AHRQ QI PQI 2013 reference population includes 29,891,024. Annual testing of rates, reliability and performance discrimination showed little change in performance between the two reference populations.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when

SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We use SDS variables for two purposes: risk adjustment and validity analyses. Because the measures are applied at the county level, we use variables that describe the make-up of the county population.

Risk adjustment: Two risk models are available; one includes age and sex make-up of the county, the other also includes the percent of households falling below the federal poverty level. These data are obtained from the US Census.

Validity analyses: In addition to the risk models we used county level demographic variables in testing. Sociodemographic variables were combined using principal component analysis (PCA) into a single socioeconomic status (SES) variable, defined at the county level. Including unemployment rate, percent adults below the federal poverty line, percent of adults aged 25-44 with some post-secondary education, percent of population not proficient in English and median household income.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Signal-to-noise

The signal-to-noise ratio refers to the entire population of US counties, comparing the degree to which rates are different from county to county (the signal) to how stable the rates are within counties (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance between counties, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences between counties (e.g. one county performs better than others).

The signal-to-noise ratio is estimated for each county. The overall signal-to-noise estimate is an average of county-level signal to noise ratios weighted by county size. County size is calculated as the eligible population for PQI 15 (population 18-39 years). Weighting by county size reduces the impact of counties that have very small denominators (the number of patients at risk).

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one county's performance from the other counties in the population, it is sensitive to the distribution of county sizes as well as the distribution of observed rates in the reference population. If the counties in a population all have performance in a narrow range, it is more difficult to reliably distinguish between counties' performance than when county performance is spread out over a much wider range. For example, if all counties have nearly perfect performance, it will be impossible to distinguish between them. As a consequence, if the distribution of county rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal to noise ratio. Different methods of calculating reliability and signal-to-noise result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8. To account for the uncertainty (noise) in a county's performance due to reliability concerns stemming from low volume, smoothed rates can be calculated.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a. Signal-to-Noise Ratio by Size Decile for PQI 15 Asthma in Younger Adults Admission Rate

Size Decile	Number of Counties	Avg. Number of Qualifying Population per County in Decile	Avg. Signal-to-Noise Ratio for Counties in Decile
1	314	1178.6	0.19649
2	314	2542	0.31188
3	314	3840.7	0.40404
4	314	5181.9	0.47145
5	314	6922.9	0.54897
6	314	9426.5	0.62095
7	314	12906.6	0.69365
8	314	19998.8	0.77833
9	314	38396.8	0.87051
10	314	191539.1	0.95615
Overall	3140	29193.4	0.7463

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 2b. SES Signal-to-Noise Ratio by Size Decile for PQI 15 Asthma in Younger Adults Admission Rate

Size Decile	Number of Counties	Avg. Number of Qualifying Population per County in Decile	Avg. Signal-to-Noise Ratio for Counties in Decile
1	314	1178.6	0.18787
2	314	2542	0.29982
3	314	3840.7	0.39039
4	314	5181.9	0.45728
5	314	6922.9	0.53481
6	314	9426.5	0.60742
7	314	12906.6	0.68140
8	314	19998.8	0.76836
9	314	38396.8	0.86400
10	314	191539.1	0.95377
Overall	3140	29193.4	0.74139

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

The indicator demonstrates good reliability with a signal-to-noise ratio of 0.75. Reliability meet threshold for all counties, except for the smallest counties with eligible populations under approximately 3800 individuals. Smoothed rates, which are recommended for all counties (and are implemented in the AHRQ software), address any reliability concerns for the smallest counties. When SES is added to the risk adjustment, the reliability remains adequate at 0.74.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e.*, *is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Systematic Assessment of Face Validity

In 2008-2009, we convened clinical panels to assess the use of the PQI across a wide range of applications, one of which was comparative reporting of area level rates.[4] We solicited nominations from national professional organization to form four clinical expert groups for a total of 73 panelists. We utilized a hybrid approach for the panel review, using two review processes, which were conducted simultaneously with information exchange between the two panels and a third and fourth panel that conducted specific reviews based on specialty. The development of this hybrid process builds from the experiences in previous panel evaluations of QI modules. The panel process that has been employed during the development of the PSIs, the PDIs and the validation of the IQIs is based on the RAND-UCLA Appropriateness Method and is termed a "nominal group" panel. The approach allowed for a wider range of input is fully described elsewhere.[5]

Panelists rated the indicator on appropriateness of use after completing a 14 item questionnaire. The questionnaire evaluated the face validity of the indicators, the panelists' perspectives on bias and potential for gaming, and the overall usefulness of the indicators when applied at one of three aspects of the health care system: area, payer and large provider organizations, for one of three purposes: internal quality improvement, comparative reporting (either public or not), and pay for performance.

Support was defined as follows:

- Full support for use: Median score of 7-9 without disagreement
- Some concern regarding use: Median score of 4-6.9 regardless of agreement status
- General support with some concerns regarding use due to disagreement: Median score of 7-9 with disagreement

---

[4] Davies SM, McDonald KM, Schmidt E, Schultz E, Geppert J, Romano, PS. 'Expanding the Uses of AHRQ's Prevention Quality Indicators: Validity from the Clinician Perspective'. *Medical Care.* 2011 Aut;49(8):479-85.

[5] Davies S, Romano PS, Schmidt EM, Schultz E, Geppert JJ, McDonald KM. Assessment of a Novel Hybrid Delphi and Nominal Groups Technique to Evaluate Quality Indicators. *Health services research.* 2011;46(6 Pt 1):2005-2018.

- Major concern regarding use: Median score or 1-3.9 regardless of agreement status

Empirical Validity

We sought to assess the relationship of county-level hospital admission rate for COPD with county level measures of socioeconomic status (SES) and community environment, heath behaviors and individual risk factors (i.e. smoking, physical activity and obesity), and access to quality care measures (i.e. primary care physician and other primary care provider density, diabetes screening testing, uninsurance). SES and community environment variables included unemployment rate, poverty rate, some college rate, English proficiency rate, median household income, food environment index, access to exercise, violent crime rate, air pollution, severe housing problems and rural status derived from the Community Health Rankings data and the American Community Survey. Because of the high number of relevant variables, we aimed to improve the interpretability of results. Principal component analysis (PCA) was used to create three composite variables for county characteristics: SES and environment (SES/E), health behaviors and personal risk factors (HB) and access to care (AC) using. The variables included in the analyses (described earlier in this paragraph) were determined a priori based on clinical and subject matter expertise. Because the relationship between the individual variables within each factor is unlikely to be consistent across all counties, we retained multiple components within each factor that explained about 70% of the variance for that factor. We also estimated prevalence based on the CDC Behavioral Risk Factor Surveillance System (BRFSS) COPD Model-based County Prevalence.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Face validity: Clinical Panel Review

During the clinical panel review both the Delphi supported the measure with some concern and Nominal panel fully supported the measure. Other topics discussed by the panel included:

- Panelists endorsed restricting the indicator to patients less than 40 years of age. Panelists felt that combining the COPD and Asthma numerator for patients 40 years and older would eliminate the diagnostic uncertainly between asthma and COPD in older patients, and thus provide a cleaner measure. COPD diagnoses in cases under 40 years of age are rare, and therefore, cases of patients less than 40 years are more likely to be true cases of asthma.
- The panel generally felt this indicator reflects issues related to access to quality outpatient care, including affordability of medication and education on proper inhaler use.
- Patient adherence to treatment recommendations remains an issue as with all chronic conditions.
- As with all chronic conditions, comorbidities and disease severity are of concern. For the asthma indicator, other respiratory conditions, infectious disease and cardiovascular conditions are of particular concern. Along with risk factors such as age, race/ethnicity, socioeconomic status, and smoking rates, panelists emphasized that environmental factors may affect admissions rates for this indicator. These environmental factors include pollution levels, altitude, allergens, housing conditions, and occupational exposures from local industries.
- Panelists also generally agreed that the high cost and complicated protocols for inhaler medications present major barriers to patient adherence to treatment recommendations. They further agreed that it is within the ability of the healthcare system to mitigate these barriers, including by providing high quality education on medication needs and inhaler use.
- Panelists felt that this indicator may also reflect some amount of “social” hospital admissions. In other words, cases in which the physician determines that social support or the home environment are insufficient for recovery outside of the hospital.
- The presence of observation units may affect admission rates.

Empirical Validity

In a negative binomial model (see section 2b2.2), prevalence, health behaviors (HB) and SES/Environment were

statistically significant predictors (p<.0001) (see Table 3a). Access to care (AC) was not significant when HB and SES/E are included in the model. However, categorizing the variables into interpretable groups (HB, AC, and SES/E) resulted in significant collinearity in the models. In particular, there was a correlation of magnitude 0.77 between IHB and one of the components for SES/E and correlations of magnitude between 0.40 and 0.50 between the components for AC and the other two components for SES/E (see Table 3b). Hence the relative importance of those factors should be interpreted with caution.

Table 3a: P-values from negative binomial model for PQI 15 as a function of prevalence and the various groups of principal components

	Prevalence	Health Behaviors	Access to Care	SES/Envr
PQI 15	0.0070	<.0001	0.31	<.0001

Table 3b: Correlation matrix of principal components (PC) used in negative binomial models.

	HB	AC_1	AC_2	SES_1	SES_2	SES_3
Health Behaviors(HB)	1.00	-0.20	-0.13	0.22	**-0.77**	0.18
Access to Care (AC_1)	-0.20	1.00	0.00	**0.50**	0.13	**0.40**
Access to Care (AC_2)	-0.13	0.00	1.00	**0.44**	0.24	0.24
Socio-Economic Status (SES_1)	0.22	-0.50	0.44	1.00	0.00	0.00
Socio-Economic Status (SES_2)	-0.77	0.13	0.24	0.00	1.00	0.00
Socio-Economic Status (SES_3)	0.18	0.40	0.24	0.00	0.00	1.00

Note that the groups of variables corresponding to health behaviors (HB) were reduced to one principal component, access to care (AC) to two, and SES to three. Correlations with magnitude larger than 0.4 are bolded.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.e., what do the results mean and what are the norms for the test conducted?)

The clinical panel noted that this indicator overall was useful. They identified specific actions that could improve rates, especially from a population health perspective, such as access to medications, patient education, reduction of risk factors, such as environmental exposure to pollution or allergens and smoking. However, it is important, as with many

population health indicators to acknowledge the complex factors influencing the indicators.

SES explained the most variance, however moderate correlation between socioeconomic status and other factors suggest that it is difficult to fully ascertain the unique contribution of each factor on asthma hospitalizations. The disparities table (see Table 2 in the supplemental files) demonstrates that zip codes in the highest income quartile have 31% lower admission rates than those in the lowest income quartile.

From a population health perspective such disparities argue for the importance of the indicator in capturing poor outcomes for vulnerable populations. Further, taking a population health perspective, efforts to decrease individual risk factors such as obesity, smoking and limited physical exercise (the variables included in the health behaviors factor) may decrease prevalence and exacerbations of asthma.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Using the 2013 data from 40 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

A total of 262 discharges were excluded due to diagnoses of cystic fibrosis and anomalies of the respiratory system. Removing this exclusion would increase the numerator count by 0.13%. The denominator does not change. Although discharges transferred into a hospital are excluded, these encounters are captured in the area level numerator via the originating hospitalization.

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The exclusion of cystic fibrosis and anomalies of the respiratory system has been retained to increase the face validity of the measure and to align the measure with the older adult and pediatric PQI/PDI for asthma/COPD admissions, although the impact on the numerator is minor for PQI15. Patients meeting the exclusion are identifiable without additional burden using the same data as is used to identify numerator qualifying discharges.

The indicator excludes patients with severe chronic respiratory diseases, because asthma in the context of these diseases differs clinically from the patients with asthma alone.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with** 23 **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Sociodemographic Factors

The risk model includes age and gender of the population.

We considered multiple income risk factors for inclusion in the model, including % of households under the federal poverty level, household income, median % of poverty level (e.g. 200% of federal poverty level). The Public Health Disparities Geocoding Project has completed extensive evaluation of alternative income variables and has demonstrated that the percent poverty variable consistently detects expected gradients in health across health outcomes, is widely available, has a low rate of missing data even at the census tract level[6]. Our team has explored alternative income variables that do not outperform the poverty level variable (data not shown).

We also considered a conceptual model that acknowledged the impact of community factors such as clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise along with community norms and beliefs. However, these are factors that are difficult to measure within the framework of the AHRQ Qis, i.e., use with administrative data. These community factors can impact prevention of asthma and self-care for asthma, which in turn can impact hospitalization rates. Poverty can be a mitigating factor, inasmuch as impoverished communities are more likely to experience housing and food insecurity[7], air pollution[8], occupational exposure and have higher smoking rates. The relationship between environment, income and health is complex and the mechanism is not fully understood.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The process to select risk factors is described in the AHRQ QI Empirical Methods report. The results of the analyses are provided in the PQI Parameter Estimates document. Both documents are available to reviewers in the supporting materials. The results of the analyses are provided in the tables below as well as on the submitted excel spreadsheet.

There are several steps involved in estimating the QI risk-adjustment models.

1. Construct candidate covariates
2. Select model covariates

---

[6] Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian S. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures-the public health disparities geocoding project. *American journal of public health.* 2003;93(10):1655-1671.

[7] Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the US. American journal of preventive medicine. 2009;36(1):74-81. e10.

[8] Bell ML, Ebisu K. Environmental inequality in exposures to airborne particulate matter components in the United States. *Environmental health perspectives.* 2012;120(2):1699-1704.

3. Estimate the models
4. Evaluate the models

Covariates are coded for each discharge record based on the data elements, data values, and logic described in the technical specifications and the appendices of the risk-adjustment coefficient tables. For a given covariate, if the discharge meets the technical specification for that covariate a value of "1" is assigned to the discharge level covariate data element. Otherwise a value of "0" is assigned to the discharge level covariate data element.

Table 4a. Risk Adjustment Coefficients, for PQI 15 Asthma in Younger Adults Admission Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCEPT		1	-8.047392	0.0178685	202829.87	<.0001
SEX	Female	1	0.837454	0.0213734	1535.2290	<.0001
AGE	Male, Age 18-24	1	-0.382994	0.0250408	233.93018	<.0001
AGE	Male, Age 25-29	1	-0.336554	0.0268152	157.52444	<.0001
AGE	Male, Age 30-34	1	-0.226800	0.0262008	74.930256	<.0001
AGE	Male, Age 35-39		Referent	.	.	.
AGE	Female, Age 18-24	1	-0.353536	0.0310396	129.72796	<.0001
AGE	Female, Age 25-29	1	-0.190547	0.0327403	33.872088	<.0001
AGE	Female, Age 30-34	1	-0.042780	0.0314820	1.8465765	0.1742
AGE	Female, Age 35-39		Referent	.	.	.
c-statistic=0.56						

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 4b. SES Risk Adjustment Coefficients, for PQI 15 Asthma in Younger Adults Admission Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCEPT		1	-8.19743673	0.025177701	106004.424	<.0001
SEX	Female	1	0.83749199	0.021373641	1535.33903	<.0001
AGE	Male, Age 18-24	1	-0.39606813	0.025049744	249.996089	<.0001
AGE	Male, Age 25-29	1	-0.34599236	0.026817633	166.453192	<.0001
AGE	Male, Age 30-34	1	-0.23108600	0.026201555	77.7845207	<.0001
AGE	Male, Age 35-39		Referent	.	.	.
AGE	Female, Age 18-24	1	-0.35522758	0.031039922	130.970085	<.0001
AGE	Female, Age 25-29	1	-0.19156186	0.03274053	34.2331361	<.0001
AGE	Female, Age 30-34	1	-0.04268203	0.031482201	1.83806134	0.1752
AGE	Female, Age 35-39		Referent	.	.	.
POVCAT	Poverty Decile 2	1	-0.00381111	0.026723409	0.02033851	0.8866
POVCAT	Poverty Decile 3	1	-0.02322692	0.02655037	0.76531867	0.3817
POVCAT	Poverty Decile 4	1	-0.01808578	0.026235351	0.47522664	0.4906
POVCAT	Poverty Decile 5	1	0.16097745	0.025270753	40.5782915	<.0001
POVCAT	Poverty Decile 6	1	0.2223969	0.024676127	81.2275687	<.0001
POVCAT	Poverty Decile 7	1	0.20873420	0.024754258	71.1029161	<.0001
POVCAT	Poverty Decile 8	1	0.14333891	0.025036027	32.7791313	<.0001
POVCAT	Poverty Decile 9	1	0.11142012	0.025997461	18.3681488	<.0001
POVCAT	Poverty Decile10 (Highest percent poverty)[1]	1	0.56469829	0.023119412	596.594715	<.0001
c-statistic=0.5453						

[1]Deciles are based on the percentage of households under the federal poverty level (FPL).
Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
See above results from validity testing in Section 2b2.3.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest (i.e. admission for asthma). The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic. The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model based on the value of the observations covariates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest

predicted probability. This creates a set of pairs of observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the non-event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common "goodness of fit" tests because these tests tend to not be informative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Table 5a. Age-sex Risk adjustment Model Discrimination and Calibration, for PQI 15 Asthma in Younger Adults Admission Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	12,349,460	0.000218	0.000228
2	9,715,818	0.000226	0.00021
3	9,514,092	0.000244	0.000235
4	8,412,305	0.000282	0.000306
5	8,817,911	0.000329	0.000302
6	11,333,912	0.000354	0.000371
7	8,192,981	0.000432	0.000397
8	8,340,546	0.00052	0.000585
9	7,664,290	0.000622	0.00055
10	7,325,882	0.000739	0.000772
C-Statistic	0.56		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 5b. Age-sex and SES Risk adjustment Model Discrimination and Calibration, for PQI 15 Asthma in Younger Adults Admission Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	10,727,468	0.000189	0.000195
2	10,919,256	0.000218	0.000207
3	10,888,415	0.000247	0.000258
4	9,955,771	0.000288	0.000281
5	8,509,855	0.000331	0.000329
6	10,480,717	0.000364	0.000374
7	8,302,290	0.000448	0.000443
8	7,625,628	0.000524	0.000525

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
9	7,726,889	0.000628	0.000602
10	6,530,908	0.000826	0.000848
C-Statistic	0.5453		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
See Table 5 in 2b4.6

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
See Table 5 in 2b4.6

2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e.*, *what do the results mean and what are the norms for the test conducted*)

A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated, as the observed to predicted values across the deciles range between 0.88– 1.12. The discrimination is low with a c-statistic of 0.56, presumably due to the limited predictors included. Addition of SES to the model slightly improves the calibration, with observed to predicted values ranging across the deciles range between 0.95 – 1.03. The c-statistic does not change substantially.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

This analysis assesses the probability that a county is higher or lower than a benchmark or threshold, given county size. It reflects whether the indicator can discriminate the best performing counties from the lower performing counties.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of counties have a lower admission rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance).

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across counties of various sizes. Each county is assumed to have an
underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each county the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is

calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Counties are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each county to determine if the county rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 6 reports the proportion of counties above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of counties not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 6a. Performance Categories by County Size Decile, for PQI 15 Asthma in Younger Adults Admission Rate

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Discharges Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	314	1,178.60	0.0000	0.0669	0.9331	0.0000	0.0000	1.0000
2	314	2,542.00	0.0000	0.1497	0.8503	0.0000	0.0127	0.9873
3	314	3,840.70	0.0000	0.2229	0.7771	0.0000	0.0159	0.9841
4	314	5,181.90	0.0000	0.1561	0.8439	0.1019	0.0064	0.8917
5	314	6,922.90	0.0000	0.3089	0.6911	0.3567	0.0350	0.6083
6	314	9,426.50	0.0000	0.2675	0.7325	0.4395	0.0255	0.5350
7	314	12,906.60	0.0000	0.3535	0.6465	0.5605	0.0287	0.4108
8	314	19,998.80	0.0000	0.4268	0.5732	0.5955	0.0541	0.3503
9	314	38,396.80	0.0000	0.5287	0.4713	0.6592	0.0510	0.2898
10	314	191,539.10	0.0637	0.7070	0.2293	0.6879	0.1656	0.1465
Overall	3,140	29,193.40	0.0064	0.3188	0.6748	0.3401	0.0395	0.6204

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

Table 6b. Performance Categories by County Size Decile, for PQI 15 Asthma in Younger Adults Admission Rate

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Discharges Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	314	1,178.60	0.0000	0.0541	0.9459	0.0000	0.0000	1.0000
2	314	2,542.00	0.0000	0.1497	0.8503	0.0000	0.0127	0.9873
3	314	3,840.70	0.0000	0.2070	0.7930	0.0000	0.0159	0.9841
4	314	5,181.90	0.0000	0.1433	0.8567	0.0892	0.0096	0.9013

			Benchmark			Threshold		
Size Decile	Number of Counties	Average Number of Denominator Discharges Per County	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
5	314	6,922.90	0.0000	0.3057	0.6943	0.3408	0.0350	0.6242
6	314	9,426.50	0.0000	0.2611	0.7389	0.4299	0.0287	0.5414
7	314	12,906.60	0.0000	0.3439	0.6561	0.5350	0.0287	0.4363
8	314	19,998.80	0.0000	0.4140	0.5860	0.5924	0.0541	0.3535
9	314	38,396.80	0.0000	0.5127	0.4873	0.6561	0.0510	0.2930
10	314	191,539.10	0.0732	0.6975	0.2293	0.6879	0.1656	0.1465
Overall	3,140	29,193.40	0.0073	0.3089	0.6838	0.3331	0.0401	0.6268

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.e., *what do the results mean in terms of statistical and meaningful differences?*)

This indicator has moderate discrimination to identify low performing counties for most counties; 38% of counties can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold). The indicator has moderate discrimination to identify high performing counties; 32% of counties can be classified as better or worse than the benchmark. Performance discrimination remains moderate when adding SES to risk adjustment.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PQI 15 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records. The rate of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

For these variables, rates of missing data are typically less than 1% of the state database. It is unlikely the bias would occur from such a low rate of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases for missing data is appropriate.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in

electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because the indicator is based on readily available administrative billing and claims data and U.S. Census data, feasibility is not an issue.
The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees. Software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2011/Methodology.html CMS Medicaid Adult Core Measures http://www.medicaid.gov/medicaid-chip-program-information/by-topics/quality-of-

	care/adult-health-care-quality-measures.html Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings Maine Health Data Organization (MHDO), MONAHRQ Website http://gateway.maine.gov/mhdo/monahrq/Methodology.html Nevada Compare Care, MONAHRQ website http://nevadacomparecare.net/ Oklahoma State Department of Health, MONAHRQ https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html Utah Department of Health, MONAHRQ website https://health.utah.gov/myhealthcare/monahrq/ Virginia Health Information, MONAHRQ website http://www.vhi.org/MONAHRQ/default.asp?yr=2013 Washington State, MONAHRQ website http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions California Office of Statewide Health Planning and Development, Healthcare Information Division http://oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/ Connecticut, Office of Health Care Access http://www.ct.gov/dph/lib/dph/ohca/publications/2010/prev_hosp_report01-2010.pdf Arkansas Department of Human Services: Arkansas Medicaid Performance http://humanservices.arkansas.gov/dms/Pages/aqg-Report-Methodology.aspx#Quality Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW) http://www.healthindicators.gov/Resources/Initiatives/CMS/Prevention-Quality-Indicators-Report_20/Indicator/Report Payment Program CMS Medicare Shared Savings Program https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html Oregon Health Authority http://www.oregon.gov/oha/analytics/Pages/CCO-Baseline-Data.aspx Regulatory and Accreditation Programs Statewide Quality Advisory Committee (Massachusetts): Center for Health Information and Analysis http://chiamass.gov/sqms/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) West Jefferson Medical Center http://www.wjmc.org/docs/WJMC-Secondary-Data-Profile-09-23-2013.pdf

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting:
Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona.
http://pub.azdhs.gov/hospital-discharge-stats/2011/Methodology.html

CMS Medicaid Adult Core Measures
http://www.medicaid.gov/medicaid-chip-program-information/by-topics/quality-of-care/adult-health-care-quality-measures.html
ACA mandated reporting system, in which Medicaid states can voluntarily report on a set of core measures.

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Connecticut.
http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine.
http://gateway.maine.gov/mhdo/monahrq/Methodology.html

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada: Quality reporting on hospitals across the state of Nevada Under NV Regulation R151-8 this transparency website presents hospital quality and utilization information.
http://nevadacomparecare.net/

Oklahoma State Department of Health, MONAHRQ
Compares quality ratings on hospitals across Oklahoma.
https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html

Utah Department of Health, MONAHRQ website
Hospital quality ratings from all hospitals in Utah.
https://health.utah.gov/myhealthcare/monahrq/

Virginia Health Information, MONAHRQ website
Compares quality ratings on hospitals across Virginia.
http://www.vhi.org/MONAHRQ/default.asp?yr=2013

Washington State, MONAHRQ website
Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals.
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions

California Office of Statewide Health Planning and Development, Healthcare Information Division
OSHPD Patient Discharge Data from all hospitals in California, totaling over 4 million records annually.
http://oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/

Connecticut, Office of Health Care Access
Preventable Hospitalizations in Connecticut: A Current Assessment of Access to Community Health Services: 2004-2009 state- and county-level hospital admission rate data from most hospitals in CT.
http://www.ct.gov/dph/lib/dph/ohca/publications/2010/prev_hosp_report01-2010.pdf

Arkansas Department of Human Services: Arkansas Medicaid Performance
Arkansas state Department of Human Services with use of Medicaid funds for children and elderly.
http://humanservices.arkansas.gov/dms/Pages/aqg-Report-Methodology.aspx#Quality

Department of Health and Human Services (DHHS), Health Indicators Warehouse (HIW)
Purpose of the HIW is to: Provide a single, user-friendly, source for national, state, and community health indicators; Facilitate harmonization of indicators across initiatives; Link indicators with evidence-based interventions.
http://www.healthindicators.gov/Resources/Initiatives/CMS/Prevention-Quality-Indicators-Report_20/Indicator/Report

Quality Improvement:
West Jefferson Medical Center
Reports indicators of potentially avoidable hospitalizations associated with the parish in which it is located, and compared those indicators with state-level indicators.
http://www.wjmc.org/docs/WJMC-Secondary-Data-Profile-09-23-2013.pdf

Regulatory/Accreditation:
Statewide Quality Advisory Committee (Massachusetts): Center for Health Information and Analysis
The committee annually recommends a standard set of health metrics to use throughout statewide health quality efforts.
http://chiamass.gov/sqms/

Payment Programs:
CMS Medicare Shared Savings Program https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html
ACO program uses quality measures to establish a performance standard to qualify for receipt of a share of savings.

Oregon Health Authority
Coordinated Care Organization (CCO) implementing Oregon´s pay for performance program using quality health metrics.
http://www.oregon.gov/oha/analytics/Pages/CCO-Baseline-Data.aspx

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
n/a

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
n/a

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

See Table 1 in response to question 1b.2.
The rate of PQI 15 hospital admissions has decreased from 2011 to 2013. In 2012 the rate was 0.5 per 1,000 while in 2013 the rate was 0.4 per 1,000. This decrease represents over 9,000 fewer hospitalizations. Further, the variation between counties decreased substantially in 2013, although we cannot determine whether this is a single-year anomaly or an actual trend. Additionally, it is important to note, the standard deviation is highly sensitive to outliers and the observed change appeared to be due to outlier counties.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
During a structured clinical panel review, panelists postulated that some uses of this indicator could disincentive care for high risk individuals. However, no evidence of this unintended consequence has arisen during actual use of the indicator. Rather, identification of high rates can help to target populations most in need of intervention.

Panelists in the same structured review and subsequent expert panel review noted that treatment of asthma in observation care may substitute for inpatient treatment, that this substitution may be systematic between areas and that this will impact the rate of the indicator. During a literature review, we identified no studies that specifically examined observation stays as a substitute for inpatient care. In a retrospective analysis of a 2002-2011 large administrative claims database of commercially insured individuals in the USA, asthma did not appear in the most frequent diagnosis categories in either emergency department-based or inpatient-based observation units.(1) A retrospective analysis of observation stays from three distinct data sources: 2010 Atlanta hospitals protocol driven observation units, 2010 Georgia hospitals for observation units (including protocol-driven, discretionary care and all bed locations), and 2009-10 National Hospital Ambulatory Medical Care Survey (NHAMCS) for similarly diverse of observation units found that asthma ranked 5th, 7th, & 8th, respectively, as the most common condition managed in observation services. However the study did not examine diagnoses by age group.(2)

1. Overman RA, Freburger JK, Assimon MM, Li X, Brookhart MA. Observation stays in administrative claims databases: underestimation of hospitalized cases. Pharmacoepidemiology and drug safety. Sep 2014;23(9):902-910.
2. Ross MA, Hockenberry JM, Mutter R, Barrett M, Wheatley M, Pitts SR. Protocol-driven emergency department observation units offer savings, shorter stays, and reduced admissions. Health Aff (Millwood). Dec 2013;32(12):2149-2156.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PQI15_NQF0283_Supplemental_Files_151214.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Carol, Stocks, Carol.Stocks@ahrq.hhs.gov
Co.3 **Measure Developer if different from Measure Steward:**
Co.4 **Point of Contact:**

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The following panelists participated in a 2009 structured panel review of the Agency for Healthcare Research and Quality Prevention Quality Indicators, which focused on evaluating expansion of the indicators to alternative denominator populations. The panel used a modified Delphi approach to evaluate the indicators, using a method that combined a nominal group technique and a Delphi technique.1 All panelists rated the indicators and received feedback from other panelists. The panelists participated in a conference call to discuss the indicators and the discussion was summarized and distributed to the group before final rating. Some panelists requested that their affiliation with this report remain anonymous, and this list is therefore a partial representation of the individuals that comprised the panels in their entirety.

1. Davies S, McDonald KM, Schmidt E, Geppert J, Romano PS. Expanding the uses of AHRQ's Prevention Quality Indicators: Validity from the clinician perspective. Med Care. Aug 2011; 49(8): 679-685.

Sandra G. Adams, MD, MS, FCCP
Pulmonolary & Critical Care Medicine
South Texas Veterans Health Care System
University of Texas Health Science Center
San Antonio, Texas
Nominated by American College of Chest Physicians

Wilbert S. Aronow, MD, FACC, FAHA, AGSF, FCCP
Geriatric Medicine, Cardiology
Westchester Medical Center
New York Medical College
Valhalla, New York
Nominated by The American Geriatrics Society

John D. Birkmeyer, MD, FACS
Surgery
University of Michigan Hospital
Ann Arbor, Michigan
Nominated by American College of Surgeons

James H. Black, III, MD
Vascular Surgery
Johns Hopkins Hospital
Johns Hopkings University
Baltimore, Maryland
Nominated by Society for Vascular Surgery

Cynthia Boyd, MD
Geriatric Medicine
Johns Hopkins Hospital, Bayview Medical Center
Johns Hopkins University
Batlimore, Maryland
Nominated by The American Geriatrics Society

Suzanne Bradley, MD
Geriatric Medicine and Infectious Diseases
Veterns Affairs Ann Arbor Healthcare System
University of Michigan
Ann Arbor, Michigan
Nominated by The American Geriatrics Society

Vickie Brown, RN, MPH, CIC
Hospital Epidemiology
University of North Carolina Hospital
Chapel Hill, North Carolina
Nominated by Association for Professionals in Infection Control and Epidemiology

John Buckley, MD, MPH
Pulmonary & Critical Care Medicine
Henry Ford Hospital
Wayne State University
Detroit, Michigan
Nominated by Society of General Internal Medicine

Pauline M. Camacho, MD, FACE
Division of Endocrinology and Metabolism
Loyola University Medical Center
Loyola Stritch School of Medicine
Maywood/Chicago, Illinois
Nominated by American Academy of Clinical Endocrinologists

Dickson Cheung, MD, MBA, MPH
Emergency Medicine
Sky Ridge Medical Center
Lone Tree, Colorado
Nominated by American College of Emergency Physicians

Leslie Davis, MSN, RN, ANP-C
Division of Cardiology
Department of Medicine
University of North Carolina
Chapel Hill, North Carolina
Nominated by American Academy of Nurse Practitioners

Barbara DeBaun, MSN, RN, CIC
Improvement Advisor
Beacon, Bay Area Patient Safety Collaborative
Bay Area Counties, California
Nominated by Association for Professionals in Infection Control and Epidemiology

Gregory J. Dehmer, MD, FACC, FSCAI
Interventional Cardiology
Scott & White Healthcare
Texas A&M University Health Science Center

Temple, Texas
Nominated by American College of Cardiology

Shawkat Dhanani, MD, MPH
Geriatric Medicine
Veterans Affairs Greater Los Angeles Healthcare System
University of California at Los Angeles
Los Angeles, California
Nominated by The American Geriatrics Society

Michelle Farber, RN, CIC
Infection Prevention and Control
Mercy Community Hospital
Coon Rapids, Minnesota
Nominated by Association for Professionals in Infection Control and Epidemiology

Amy Fendrich, MD
Internal Medicine
Memorial Regional Hospital Primary Care Clinics
Dania Beach, Florida
Nominated by American Puclic Health Association, Medical Care Section

Carlos M. Ferrario, MD, FAHA, FASA, FACC
Cardiology
Wake Forest University Health Science Center
Winston-Salem, North Carolina
Nominated by American College of Cardiology

John E. Gardella, MD, MBA, FCCP, FHM
Hospital Medicine
Presbyterian Hospital, Novant Health
Charlotte, North Carolina
Nominated by Society of Hospital Medicine

Eric Gertner, MD, MPH, FACP
Internal Medicine
Lehigh Valley Hospital
Penn State College of Medicine
Allentown, Pennsylvania
Nominated by American College of Physicians

James M. Gill, MD, MPH
Family Medicine
Delaware Valley Outcomes Research
Jefferson Medical College
Wilmington, Delaware
Nominated by American Puclic Health Association, Medical Care Section

Louis Gilleran, MD, MPH, FACPM
Preventive Aerospace Medicine
Naval Medical Center San Diego
San Diego, California
Nominated by American Puclic Health Association, Medical Care Section

Christopher Gonzalez, MD, MBA
Urology
Northwestern University Medical Center

Chicago, Ilinois
Nominated by American Urological Association

Michael K. Gould, MD, MS
Pulmonary and Critical Care Medicine
Veterans Affairs Palo Alto Health Care System
Stanford University
Palo Alto, California
Nominated by American Thoracic Society

George Grunberger, MD, FACP, FACE
Diabetes, Endocrinology
Grunberger Diabetes Institute
Wayne State University
Bloomfield Hills & Detroit, Michigan
Nominated by American Academy of Clinical Endocrinologists

A. Seiji Hayashi, MD, MPH
Family Medicine
Unity Health Care, Inc.
George Washington University
Washinton, DC
Nominated by American Puclic Health Association, Medical Care Section

Richard Hellman, MD, FACP, FACE
Clinical Endocrinology
North Kansas City Hospital
University of Missouri, Kansas City School of Medicine
Kansas City, Missouri
Nominated by American Academy of Clinical Endocrinologists

Steve J. Hodges, MD
Surgery, Pediatric Urology
Wake Forest University Health Sciences
Institute for Regenerative Medicine
Winston-Salem, North Carolina
Nominated by American College of Surgeons (by proxy through Dr. Anthony Atala)

Mary Johnson, MS, RD, CDE, BC-ADM
Diabetes Quality and Education
Geisinger Health System
Central Pennsylvania
Nominated by American Dietetic Association

Jeanette Kalupa, MSN, ACNP-BC, APNP
Acute Care Nurse Practioner, Hospitalist
Cogent Healthcare
Aurora St. Luke´s Medical Center
Milwaukee, Wisconsin
Nominated by American Academy of Nurse Practitioners

Marjorie L. King, MD, FACC, FAACVPR
Cardiology
Helen Hayes Hospital
Columbia University
West Haverstraw, New York
Nominated by American College of Cardiology

Jerry A. Krishnan, MD, PhD
Pulmonary and Critical Care Medicine
University of Chicago, Asthma and COPD Center
Chicago, Ilinois
Nominated by American Thoracic Society

Geoffrey Lamb, MD
Internal Medicine
Froedtert Hospital
Medical College of Wisconsin
Milwaukee, Wisconsin
Nominated by Society of General Internal Medicine

Gene Lambert, MD, MBA
Hospital Medicine
Massachusetts General Hospital
Harvard University
Boston, Massachusetts
Nominated by Society of Hospital Medicine

David E. Lanfear, MD, MS, FACC
Cardiology
Henry Ford Hospital
Wayne State University
Detroit, Michigan
Nominated by American College of Cardiology

Karen Lasser, MD, MPH
Primary Care, Internal Medcine
Boston University, Medical Center
Boston, Massachusetts
Nominated by Society of General Internal Medicine

Susan Lee, MS, NP-C, FAANP
Family Medicine, Internal Medicine
Bloomington Hospital
Bloomington, Indiana
Nominated by American Academy of Nurse Practitioners

Edgar V. Lerma, MD
Nephrology
Associates in Nephrology, S.C.
University of Illinois at Chicago
Chicago, Illinois
Nominated by The American Society of Nephrology

James T. Li, MD, PhD
Internal Medicine, Allergy
Mayo Clinic, College of Medicine
Rochester, Minnesota
Nominated by American Academy of Allergy Asthma and Immunology

Jenny J. Lin, MD
Internal Medicine
Mount Sinai Medical Center
Mount Sinai Medical School

New York, New York
Nominated by Society of General Internal Medicine

Frank LoGerfo, MD
Vascular Surgery
Beth Israel Deaconess Hospital
Harvard Medical School
Boston, Massachusetts
Nominated by Society for Vascular Surgery

John J. Lopez, MD, FACC
Interventional Cardiology
Loyola University Medical Center
Stritch School of Medicine, Loyola University
Maywood, Illinois
Nominated by American College of Cardiology

Thomas D. MacKenzie, MD, MSPH
Internal Medicine
Denver Health
University of Colorado
Denver, Colorado
Nominated by Society of General Internal Medicine

William Marston, MD
Vascular Surgery
University of North Carolina
Chapel Hill, North Carolina
Nominated by Society for Vascular Surgery

Helmut Meisl, MD, FACEP
Emergency Medicine
Good Samaritan Hospital
San Jose, California
Nominated by American College of Emergency Physicians

Mark L. Metersky, MD, FCCP
Pulmonary and Critical Care Medicine
University of Connecticut, Health Center
Farmington, Connecticut
Nominated by American College of Chest Physicians

Lena M. Napolitano, MD, FACS, FCCM
Surgery
University of Michigan
Ann Arbor, Michigan
Nominated by American College of Surgeons

John A. Parker, Jr., MD
Family Medicine
Marshall University
Huntington, West Virginia
Nominated by American Academy of Family Physicians

Jay I. Peters, MD, FCCP
Chief Pulmonary & Critical Care Medicine
University of Texas Health Science Center

San Antonio, Texas
Nominated by American College of Chest Physicians

Michael P. Phelan, MD, FACEP
Emergency Medicine
Cleveland Clinic
Cleveland, Ohio
Nominated by American College of Emergency Physicians

Mark Potter, MD
Family Medicine
University of Illinois at Chicago, Medical Center
Chicago, Illinios
Nominated by American Academy of Family Physicians

Carol Rasmussen, MSN, NP-C, CDE
Diabetes Program
Exodus Healthcare
Magna, Utah
Nominated by American Association of Diabetes Educators

H. David Reines, MD, FACS
Surgery
Inova Fairfax Hospital
Virginia Commonwealth University
Falls Church, Virginia
Nominated by American College of Surgeons

Barry Saver, MD, MPH
Family Medicine
University of Massachusetts Memorial
Worcester, Massachusetts
Nominated by American Academy of Family Physicians

Douglas J.E. Schuerer, MD, FACS, FCCM
Acute and Critical Care Surgery
Barnes Jewish Hospital
Washington University
St. Louis, Missouri
Nominated by American College of Surgeons

June Schulz, RRT, FAACVPR
Respiratory Care, Pulmonary Rehabilitation
Sanford University of South Dakota Medical Center
University of South Dakota
Sioux Falls, South Dakota
Nominated by American Association of Cardiovascular and Pulmonary Rehabilitation

Kristine M. Thompson, MD
Emergency Medicine
Mayo Clinic
Jacksonville, Florida
Nominated by American College of Emergency Physicians

Francesca J. Torriani, MD, FIDSA
Infectious Disease, Epidemiology
University of California at San Diego

San Diego, California
Nominated by Infectious Disease Association of California

Dace L. Trence, MD, FACE
Division of Metabolism, Endocrinology, and Nutrition
University of Washington Medical Center
Seattle, Washington
Nominated by American Academy of Clinical Endocrinologists

Arjun K. Venkatesh, MD, MBA
Emergency Medicine
Brigham and Women´s Hospital, Massachusetts General Hospital
Harvard University
Boston, Massachusetts
Nominated by American College of Emergency Physicians

Raoul Wolf, MBBCh, FCCP, FAAAI
Allergy and Clinical Immunology
La Rabida Chidrens Hospital, Comer
University of Chicago
Chicago, Illinois
Nominated by American College of Chest Physicians

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2007

Ad.3 **Month and Year of most recent revision:** 11, 2007

Ad.4 **What is your frequency for review/update of this measure?** Annually

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.6 **Copyright statement:** The AHRQ QI software is publicly available. We have no copyright disclaimers.

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0334
De.2. **Measure Title:** PICU Severity-adjusted Length of Stay
Co.1.1. **Measure Steward:** Virtual PICU Systems, LLC
De.3. **Brief Description of Measure:** The number of days between PICU admission and PICU discharge.
1b.1. **Developer Rationale:** Use of this measure, coupled with 0335 as a balancing measure, allows for evaluation of appropriateness of resource utilization.

Literature review:
Adult data has found 24% of admissions to adult ICUs were for observation only in one study (1), while 77% of admissions to an adult ICU were for monitoring alone in another study (2).

Pediatric critical care studies found that 27% of admissions to one PICU received no benefit beyond what could be provided elsewhere (3), while an analysis of eight PICUs demonstrated varying efficiency with a potential saving of 5.1 to 17.2% of ICU days of care through earlier discharge (4).

S.4. **Numerator Statement:** Number of PICU days, PICU days = Number of days between PICU admission and PICU discharge.(For all eligible patients admitted to the ICU, the time at discharge from ICU minus the time of ICU admission (first recorded vital sign on ICU flow sheet)
S.7. **Denominator Statement:** The denominator is the average (mean) predicted length of stay using the adjustment model.
S.10. **Denominator Exclusions:** Patients => 18 years of age

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims, Electronic Clinical Data : Registry, Paper Medical Records
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Jul 31, 2012

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Recommend use in conjunction with 0334 (PICU Severity Adjusted Length of Stay) and 0335 (PICU Unplanned Readmission Rate) as balancing measures.

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This outcome measure indirectly measures process (decision making related to PICU discharge) while directly measuring PICU resource utilization.
- Pediatric critical care studies found that 27% of admissions to one PICU received no benefit beyond what could be provided elsewhere, while an analysis of eight PICUs demonstrated varying efficiency with a potential saving of 5.1 to 17.2% of ICU days of care through earlier discharge.
- The developer states measurement of ICU and PICU length of stay (LOS) was included as a measure or focus of study in more than 9,000 publications in a recent Pubmed search. Risk-adjustment of LOS is an established and accepted methodology.
- The developer recommends this measure be paired with 0335, its unplanned readmission measure.
- The previous Committee concluded the evidence demonstrates the importance of a risk adjustment model and supports the use of a LOS metric.

Guidance from the Evidence Algorithm: 1→2 (eligible for PASS rating)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale for this outcome measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

- Current performance data are derived from 79 PICUs submitting PRISM III data for patients discharged in 2014; 92,928 individual PICU discharges and transfers were recorded.
  - The unit-level Severity-adjusted Length of Stay Ratio (SLOSR) ranges between 0.66 and 1.822.
  - For 2014, the median unit-level SLOSR was 1.01, with interquartile rate of 0.89-1.14.

- The mean unit-level SMR was 1.03 with standard deviation of 0.20. The patient-level mean SLOSR for 2014 is 0.986 (95% CI, 0.978-0.995).

- For performance over time: An analysis of 273,130 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e., no increasing or decreasing trend) for SLOSR by quarter (Spearman's correlation test p=0.10).
- Pediatric critical care studies found that 27% of admissions to one PICU received no benefit beyond what could be provided elsewhere, while an analysis of eight PICUs demonstrated varying efficiency with a potential saving of 5.1 to 17.2% of ICU days of care through earlier discharge.

Disparities

The developer provides the following:

- Historically, population differences have not been found to be variable in pediatric intensive care therapies. One study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status.
    - After adjustment for differences in illness severity, standardized mortality ratios and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care unit.
    - Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.
- VPS stratifies by race/ethnicity, age groups, gender, and insurance payer. While the developer determined statistically significant differences, it emphasizes that whether these are clinically relevant differences is unclear. The developer also notes any differences may reflect prehospitalization factors independent of the care provided in the ICU. Specific results are as follows:
    - There was a statistically significant difference for SLOSR between Males (0.97; 95% CI, 0.96–0.98), and Females (1.01; 95% CI, 1.00–1.02).
    - Except for two categories, no statistically significant differences were observed for the age categories ('<1 Month', '1 Month - 23 Month', '2 Years - 5 Years', '6 Years - 12 Years', '13 Years - 18 Years'). First, the age group '<1 month' (n=3,379) had a SLOSR statistical lower than the other age categories (0.92; 95% CI, 0.89-0.95, p=0.0008). Second, the age group '6 Years - 12 Years' (n=21,201) had a SLOSR statistical higher than other age categories (1.02; 95% CI, 1.00-1.04, p=0.0003).
    - Three races/ethnicities had a statistical higher SLOSR than the other groups: ´Asian/Indian/Pacific Islander´ (n=2,958) (1.05; 95% CI, 1.00-1.10, p=0.020), ´Hispanic´ (n=12,615) (1.03; 95% CI, 1.01-1.05, p=0.0003), and ´Other/Mixed´ (n=4,407) (1.07; 95% CI, 1.03 – 1.11, p=0.0003). One races/ethnicity had a statistical lower SLOSR than the groups: ´African American´ (n=17,309) (0.95; 95% CI, 0.93-0.97, p<0.0001). Two groups did not have a statistical difference of SLOSR 'Caucasian/European Non-Hispanic' and 'American Indian/ Indigenous'.
    - Two insurance categories had a statistical higher SLOSR than the categories: ´Managed Care´ (n= 5,835) (1.06; 95% CI, 1.03-1.10, p=0.019), and ´Medicaid & Medicaid Managed Care´ (n=19,311) (1.04; 95% CI, 1.03-1.06, p=0.0008). Two insurance categories had a statistical lower SLOSR than the categories: ´Commercial/Indemnity Insurance´ (n=6,878) (0.98; 95% CI, 0.95-1.00, p=0.0002), and Self-pay (n=739) (0.80; 95% CI, 0.73-0.87, p=<0.0001). One insurance category did not have a statistical different SLOSR than the

categories: Other lower (n=1,483) (0.96; 95% CI, 0.90-1.02, p=0.051). Primary payer is reliably collected by 38% of sites.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The developer indicates that this is an outcomes measure. The measure does relate to the average length of stay, with a proprietary risk adjustment model that calculated using a statistical model. The outcome measured could be tied back to at least one healthcare action including processes, interventions and services.

**This is an outcome measure. The stated rationale does not define any specific structure, process, intervention, or service that may be related to the outcome. The measure focuses on LOS and may demonstrate differences in resource use between facilities. There are many processes of care within a PICU that may impact LOS.

**Reasonable evidence with same caveats listed for other VPS measures. Proprietary and old SOI adjustment

1b. Performance Gap

Comments:

**The developer identified statistical differences in performance based on race/ethnicity. However, the developer could not determine if the differences were due to care provided in the PICU or if they may have been a result of factors prior to admission to the PICU. Evidence provided did not clearly indicate that there were any differences in outcomes due to the service delivery or length of stay in the PICU for varying sub populations.

**The severity-adjusted length of stay ratio varied significantly between PICUs (range 0.66-1.822). I did find it troubling that despite use from 2012 through 2014, the developers were able to demonstrate no trends on the measure.

The developers stratified performance on the measure by race/ethnicity, age groups, gender, and insurance payer and found no clinically significant differences.

**Performance gap very likely exists

1c. High Priority (previously referred to as High Impact)

Comments:

**This is not a composite measure though the developer did recommend using it in conjunction with other measures that utilized its proprietary system.

**Not applicable

**na

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data : Registry, Paper Medical Records

Specifications:

- The developer reports the specifications were updated since the last review, as follows:
  - The developer updated the original PRISM length of stay (LOS) model by adding more predictors and re-estimating the coefficients. It developed the linear regression model for LOS on the training data set (based on admissions between Q2 2009 and Q1 2013, n=275,013), and independently confirmed the performance of the resulting model on the validation dataset (based on admissions between Q2 2013 and Q1 2014, n=73,705).
  - Several changes to predictor variables were made from the previous endorsement:
    - The 'PRISM3 Score' is considered to have a non-linear (i.e. quadratic) relationship with the outcome. Previously, PRISM3 Score assumed a piece-wise linear relationship.
    - The 'age categories' are slightly different (using only neonate (<1m) and infant (1m-1y)) than before (neonate, infant, 1y-3y, 3y-6y, 6y-12y, >12y). The 'age categories' other than neonate and infant where not statistically significant in revised regression model, and found to be 'marginally' statistically significant in the previous regression. While the additional age categories seemed reasonable to include, the effect on PICU LOS for other age groups can be adjusted by other predictors within the model. Therefore, to simplify the model and reduce noise, the developer removed the additional 'age categories' not statistically significant.
    - The predictor variable ('Is patient associated with an acute problem?') was added to the 2014 regression model. In the 2014 revision, the variable has a statistical effect in the model ($p<0.0001$) and therefore remained in the regression. In the previous regression model however, it was found to have a non-statistical effect and removed from the previous model.
    - One predictor variable ('Did the patient experience head trauma?') was used in the previous regression model, however in 2014 the developer found it to have a non-statistical effect and removed it from the final 2014 model.
- The numerator is: *The number of PICU days. PICU days are defined as number of days between PICU admission and PICU discharge. For all eligible patients admitted to the ICU, the time at discharge from ICU minus the time of ICU admission (first recorded vital sign on ICU flow sheet)*. The denominator is: *The average (mean) predicted length of stay using the adjustment model.*
- The calculation algorithm is stated in S.18.

Questions for the Committee :

- *Are the specification changes appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The measure used the PRISM III algorithm, a proprietary risk adjustment scheme.

- Information about the model has been published previously.

Describe any updates to testing: Initial and quarterly interrater reliability from all clinical data collectors for each unit participating in VPS was calculated.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- The developer conducted validity testing at the data element level. Per NQF guidance, separate reliability testing is not required when validity testing at the data element level is performed for all critical data elements.

Results of reliability testing:

- The results of the data element level validity testing are provided in the following section.

Guidance from the Reliability Algorithm : Not applicable

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The measure used the PRISM III algorithm, a proprietary risk adjustment scheme.
- Information about the model has been published previously.

Describe any updates to validity testing

- The developer updated data element-level validity testing. Additionally, to assess the predictive model's validation, performance was evaluated in four ways using the validation data set.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

The developer provides the following information:

- VPS requires initial and quarterly interrater reliability from all clinical data collectors for each unit participating in VPS. The developer does not explicitly indicate all critical data elements are assessed during this process.
- A dataset of 348,718 PICU admissions was split into a training data set (n=275,013) based on admissions between Q2 2009 and Q1 2013, and an independent dataset for validation (n=73,705) based on admissions between Q2 2013 and Q1 2014.
- Validation of the model/performance score involved four analyses performed on the validation set:
    - A paired t-test was used to compare the mean observed ICU LOS to the mean predicted ICU LOS for the entire validation population and for various age categories and other subgroups.
    - Coefficients of determination were calculated to measure the variance in LOS by patient and by unit. A regression of the mean observed LOS against the mean predicted LOS was used to assess the proportion of variation across patients (and again for units) which is explained by the model.
    - Calibration curves comparing the average expected LOS to the average observed LOS by dividing the validation data set into 10 equal sized cohorts (i.e., deciles of predicted LOS)
    - Evaluating the model calibration by checking that the distribution of the PICUs' SLOSRs

Validity testing results:

The developer reports the following results:

- For the data element-level validity testing, the 2014 aggregate IRR concordance is 96.81%. The developer does not provide additional analyses (e.g., sensitivity, specificity, positive predictive value, negative predictive value).
- The overall t-test shows that the mean difference between the observed LOS and the predicted LOS is not statistically different (6.7 minutes; p=0.79). For the age categories, no statistical significant differences were found for age the groups '<1 months', '1 months-23 months', '2 years-5 years', '6 years-12 years', '12 years-18 years'.
- The predictive model had an R-squared of 0.23 at the patient-level. At the ICU level, the predictive LOS model had an R-squared of 0.92.
- The calibration curve shows good model fit with R-square of 0.94 (Figure 1).
- The PICU-level SLOSRs have a normal distribution centered at 1.0 (p=0.15) (Figure 2).

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following information:

- The endorsed severity adjustment methodology used for calculating severity adjusted LOS excludes:
    - PICU patients >=18 years of age
    - PICU patients under the age of 18 years with a stay < 2 hours in the PICU or < 2 consecutive sets of vital signs consistent with life
    - Patients admitted to PICU for palliative care
        - The developer states palliative cases are excluded because the intention is that the patient will likely die in the ICU, skewing SMR calculations if the patients are included.
- The developer states the other exclusions are consistent with the use of the PRISM 3 instrument for severity adjustment. The tool has not been validated in patients <36 weeks gestation, > or equal to 18 years, or if not in the PICU at least two hours/for two vital signs to be taken.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

The developer reports the following about its risk model:

- The developer posits four selection criteria for risk adjustment tool for pediatric ICUs; it indicates its model meets these criteria:
    - Tool must allow quality assessment and comparison between intensive care units, and must be widely used
    - Tool must be valid and reliable for severity adjustment and measurement of quality of care provided
    - Computation of mortality risk must be in the public domain (i.e., free of charge)
    - Algorithms must receive ongoing validation and recalibration
- Since the latest model release is intended to be a refresh of the PRISM III LOS model, the developer used predictors that are included in PRISM III Risk of Mortality (ROM) and did not include interaction terms or site level predictors. The LOS (in days) is predicted from the following terms at the patient-level:
    - PRISM3 Score
    - Neonatal (less than 1 month) patient,
    - Infant (1 month to 1 year) patient,
    - Post-operative patient,
    - Admission of patient from Inpatient Unit,
    - Previous ICU admission,
    - Patient with an oncology diagnosis,
    - Patient with an acute overdose,
    - Patient with acute diabetes,

- Patient with an operative cardiac disease,
- Patient with pneumonia,
- Patient with non-head trauma,
- Patient associated with an acute problem, and
- Patient on mechanical ventilation.

- In 2014, the PRISM 3 predicted length of stay model developed by Pollack was recalibrated by VPS.
  - More than 240,000 PICU admissions were used to develop a linear regression model to estimate physical ICU length of stay. ICU length of stay values were truncated at 30 days.
  - The regression model used 15 predictive variables captured within the first 12 hours of ICU admission, including the PRISM 3 score. The developer states the model was extensively validated with an independent dataset of more than 81,000 cases. The 2014 predicted length of stay model outperforms the 2006 model.
- The standardized length of stay ratio (SLOSR) is created by dividing the sum of observed physical length of stay (truncated at 30 days) by the sum of predicted length of stay. Cases must meet PRISM 3 inclusion criteria to receive a PRISM 3 length of stay prediction.
- The developer states it provides information to participating ICUs that are stratified by race/ethnicity and insurance payer, but the model does not appear to include these risk factors. While the developer found statistically significant differences based on race/ethnicity and insurance payer, it emphasizes that whether these are clinically relevant differences is unclear.
- The developer also notes any differences may reflect prehospitalization factors independent of the care provided in the ICU.
-

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Does the Committee concur with the developer's rationale for not including SDS factors in the risk model?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer states the data available from the VPS system reveals that the severity adjusted length of stay among 80 participating PICUs ranged from 1.71 to 4.02 days in the third quarter of 2011, indicating unit-specific variance. The developer concludes that since numerators, denominators, and all definitions are standardized with an IRR >96%, this variation reflects differences in care and not the measurement itself.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- Missing data analysis was not conducted on this measure

Guidance from the Validity Algorithm : 1→2→4→5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The specifications for the measure are consistent with the evidence. The target population is clearly defined and the results could be meaningful if they point to opportunities for care improvement. More work would need to be done to get to causal factors in order to develop appropriate interventions to implement effective interventions.
**The model was validated by taking a large PICU data set which was divided into a derivation data set and a validation data set. I did not find evidence that the statistical model was compared to any other clinical models (e.g., targeted case review).
**reasonable

2a2. Reliability Testing

Comments:

**The validity testing had an adequate number of patients in order to calculate results. Based on the results it appears that the measure is valid and could be generalized and allow for conclusions to be drawn regarding the quality of care/service in the PICU.
**Validity testing was consistent with other statistical models to evaluate risk-adjusted differences in patient outcomes.
**testing appears adequate and more robust than most

2b2. Validity Testing

Comments:

**Administrative data was used in this measure so there were no concerns with missing data or no response data. Of concern though is the developer's statement "The developer was unable to collect comparison data for measuring improvement because of recalibration of the tool. It reports improvement is now being tracked anew, and the data will be available next year."
**I did not find a description of how missing data is treated in the statistical model.
**do design threats identifiable

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias

Comments:

**Reliability testing was adequate and at the data element level. It was updated since its last review.
**The developers report inter-rater reliability testing of 96.8% concordance for those units participating in the registry.
**differences in performance can be identified

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer provides the following information:

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. All necessary data may be available

electronically if an organization has implemented an EHR. In the absence of an EHR, the developer notes that manual data collection is required.

- There are no fees or licensing required to collect and calculate this measure. However, VPS provides these data free of charge to participating member units. The 2016 VPS participation fee schedule consists of a baseline fee with incremental charges based on unit volume. Small units pay $16,590 per annum and the largest units pay $33,150 per annum; the 2016 VPS fee schedule attached in appendix A1.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments:

**This measure is feasible and the data elements necessary to calculate the rate should be contained in the medical record and in an electronic health record. There are no concerns with the data collection strategy. It could be easily operationalized.

**Most of the data elements should be available in an electronic format. The PRISM III score may need to be calculated but I assume this is built into the registry program. Claims data and EHR data should provide most of the data elements needed.

**Proprietary

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- The measure data is not aggregated and publicly reported; however, some hospitals participating in the VPS system may individually publicly report their data.
- The measure is publically reported at the Texas Children Hospital
- The measure is part of a payment program at the California Children Health Services, but the developer does not indicate if public reporting is a component of this program.
- VPS currently provides data to 128 U.S. PICUs/CICUs from 116 hospitals, which use the measure to benchmark against a VPS reference group of other participants.

Improvement results:

- The developer was unable to collect comparison data for measuring improvement because of recalibration of the tool. It reports improvement is now being tracked anew, and the data will be available next year.

Unexpected findings (positive or negative) during implementation: The developer states there were no unexpected findings during implementation.

Potential harms: The developer reports there were no identified unintended consequences for this measure during testing or since implementation.

Feedback: No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Is use only in local programs sufficient? Should the measure be publicly reported on a broader scale?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Limited public reporting is occurring with this measure. It is being publicly reported by two hospitals. 116 hospitals use the system and use system data to benchmark against the performance of other hospitals.
**There is very limited use of the measure for public accountability. The measure is apparently used by the California Childrens & Acute Health Services unit for benchmarking and payments. The principal use of the measure is for internal quality improvement efforts and benchmarking between PICUs that submit data to VPS.
**How does public reporting impact our decision making here?

Criterion 5: Related and Competing Measures

Related or competing measures

- 0335: PICU Unplanned Readmission Rate
- 0702: Intensive Care Unit (ICU) Length-of-Stay (LOS)

Harmonization

- Developer indicates this measure and 0335 have been paired. The measure specifications are harmonized with NQF 0335.
- Developer did not provide harmonization information for 0702.

NQF #: 0334 NQF Project: Pulmonary Project

1. IMPACT, OPPORTUITY, 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)

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 indirectly measures process (decision making related to PICU discharge) while directly measuring PICU resource utilization.

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

Selected individual studies (rather than entire body of evidence)

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)*:

Measurement of ICU and PICU length of stay (LOS) was included as a measure or focus of study in over 9000 publications in a recent Pubmed search. Risk-adjustment of LOS is an established and accepted methodology.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*:

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):* Length of stay may vary significantly based on the severity of illness of the patient at the time of admission. Failure to adjust for patient-level severity of illness may result in inappropriate comparison.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction*

of the effect): There are two categories of studies measuring LOS: those that use a risk-adjustment method and those that don´t. For those studies that compare the two approaches, risk adjustment is found to be an important factor. However, short of the studies that compare methodologies, there is no way to otherwise make comparisons

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

Risk-adjustment of LOS accounts for otherwise unexplained variation within and between centers which may result in flawed interpretations of performance.

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:

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

1c.12 If other, identify and describe the grading scale with definitions: per 1c.9 above no grading has been done.

1c.13 Grade Assigned to the Body of Evidence:

1c.14 Summary of Controversy/Contradictory Evidence: None

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

Ruttimann UE, Pollack MM. Variability in duration of stay in pediatric intensive care units: A multiinstitutional study. The

Journal of Pediatrics;1996:128(1), 35-43.

Straney L, Clements A, Slater A. Quantifying variation of paediatric length fo stay among intensive care units in Australia and New Zealand. Qual Saf Health Care. 2010. 19 1-5

Starney LD, Clement A, Alexander J, Slater A. Measuring efficiency in Australian and New Zealand paediatric intensive care units. Intensive Care Med. 2010; 36(8):1410-16

Niskanen M, Reinikainen M, Pettila V. Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. Intensive Care Med 2009; 35(6). 1060-7

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

N/A

1c.17 Clinical Practice Guideline Citation:

1c.18 National Guideline Clearinghouse or other URL: N/A

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: per 1c.19 above no grading has been done.

1c.23 Grade Assigned to the Recommendation:

1c.24 Rationale for Using this Guideline Over Others:

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: Moderate 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0334_Evidence_MSF5.0_Data.123015..doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

Use of this measure, coupled with 0335 as a balancing measure, allows for evaluation of appropriateness of resource utilization.

Literature review:

Adult data has found 24% of admissions to adult ICUs were for observation only in one study (1), while 77% of admissions to an adult ICU were for monitoring alone in another study (2).

Pediatric critical care studies found that 27% of admissions to one PICU received no benefit beyond what could be provided elsewhere (3), while an analysis of eight PICUs demonstrated varying efficiency with a potential saving of 5.1 to 17.2% of ICU days of care through earlier discharge (4).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Current performance:

The data from 79 PICUs submitting PRISM III data for patients discharged in 2014. In total, 92,928 individual PICU discharges and transfers were recorded.

The unit-level Severity-adjusted Length of Stay Ratio (SLOSR) ranges between 0.66 and 1.822. Over 2014, the median unit-level SLOSR was 1.01 with interquartile rate of 0.89 – 1.14. The mean unit-level SMR was 1.03 with standard deviation of 0.20.

The patient-level mean SLOSR for 2014 is 0.986 (95% CI, 0.978 – 0.995).

Performance overtime:

An analysis of 273,130 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e. no increasing or decreasing trend) for SLOSR by quarter (Spearman's correlation test p=0.10).

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**

1.Knaus WA, Draper EA, Wagner DP: The use of intensive care: New research initiatives and their implications for national health policy. Milbank Q
1983;61:561-583

2. Thibault GE, Mulley AG, Barnett CO, et al: Medical intensive care: Indications, interventions, and outcome. N Engl J Med 1980;302:938-942.

3. Pollack MM, Ruttimann UE, Glass NL, et al: Monitoring patients in pediatric intensive care. Pediatrics 1985;76:719-724.

4. Pollack MM. Getston PR, Ruttimann UE, et al. Efficiency of intensive care. A compartive analysis of eight pediatric intensive care units. JAMA 1987; 358:1481-1486

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients;*

dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.

Historically, population differences have not been found to be variable in pediatric intensive care therapies. A study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status. After adjustment for differences in illness severity, standardized mortality ratios and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care unit. Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.

VPS stratifies by race/ethnicity, age groups, gender, and insurance payer.

There was a statistically significant difference for SLOSR between Males (0.97; 95% CI, 0.96–0.98), and Females (1.01; 95% CI, 1.00–1.02).

Except for two categories, no statistically significant differences were observed for the age categories ('<1 Month', '1 Month - 23 Month', '2 Years - 5 Years', '6 Years - 12 Years', '13 Years - 18 Years'). First, the age group '<1 month' (n=3,379) had a SLOSR statistical lower than the other age categories (0.92; 95% CI, 0.89–0.95, p=0.0008). Second, the age group '6 Years - 12 Years' (n=21,201) had a SLOSR statistical higher than other age categories (1.02; 95% CI, 1.00 – 1.04, p=0.0003).

Three races/ethnicities had a statistical higher SLOSR than the other groups: ´Asian/Indian/Pacific Islander´ (n=2,958) (1.05; 95% CI, 1.00 – 1.10, p=0.020), ´Hispanic´ (n=12,615) (1.03; 95% CI, 1.01 – 1.05, p=0.0003), and ´Other/Mixed´ (n=4,407) (1.07; 95% CI, 1.03 – 1.11, p=0.0003). One races/ethnicity had a statistical lower SLOSR than the groups: ´African American´ (n=17,309) (0.95; 95% CI, 0.93 – 0.97, p<0.0001). Two groups did not have a statistical difference of SLOSR 'Caucasian/European Non-Hispanic' and 'American Indian/ Indigenous'.

Two insurance categories had a statistical higher SLOSR than the categories: ´Managed Care´ (n= 5,835) (1.06; 95% CI, 1.03 – 1.10, p=0.019), and ´Medicaid & Medicaid Managed Care´ (n=19,311) (1.04; 95% CI, 1.03 – 1.06, p=0.0008). Two insurance categories had a statistical lower SLOSR than the categories: ´Commercial/Indemnity Insurance´ (n=6,878) (0.98; 95% CI, 0.95 – 1.00, p=0.0002), and Self-pay (n=739) (0.80; 95% CI, 0.73 – 0.87, p=<0.0001). One insurance category did not have a statistical different SLOSR than the categories: Other lower (n=1,483) (0.96; 95% CI, 0.90 – 1.02, p=0.051). Primary payer is reliably collected by 38% of sites.

While we determined statistically significant differences it is not clear that these are clinically relevant differences. Further, any differences that exist may reflect prehospitalization factors independent of the care provided in the ICU.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Lopez A, Tilford J, Anand K, et al. Variation in pediatric intensive care therapies and outcomes by race, gender and insurance status. Ped Crit Care Med 2006; 7(1). 2-6.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of

patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
ICUs are a source of significant health care cost (1,2,5) PICUs have been shown to have varying degrees of efficiency with consumption of resources that could have been provided elsewhere (3,4). This, coupled with the potential for hospital acquired infections, supports caring only for those patients in the ICU who require ICU-level care.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Russell LB: The role of technology assessment in cost control, in McNeil BJ, Cravalho EG (eds): Critical Issues in Medical Technology. Boston, Auburn House, 1980, pp 129-138.
2. Knaus WA, Draper EA, Wagner DP: The use of intensive care: New research initiatives and their implications for national health policy. Milbank Q
1983;61:561-583
3. Pollack MM, Ruttimann UE, Glass NL, et al: Monitoring patients in pediatric intensive care. Pediatrics 1985;76:719-724.
4. Pollack MM. Getston PR, Ruttimann UE, et al. Efficiency of intensive care. A compartive analysis of eight pediatric intensive care units. JAMA 1987; 358:1481-1486
5. Brilli RJ, Spvetz A, Branson, RD, et al. Critical care delivery in the intensive care unit: Defining clinical roles and the best practice model. Critical Care Medicine; 2001: 29 (10), 2007-2019.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety : Readmissions

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://s3.amazonaws.com/vpspublic/NQFMeasures.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since last endorsement cycle the predicted length of stay algorithm has been re-calibrated with improved performance. VPS has updated the original PRISM LOS model by adding more predictors and re-estimating the coefficients. We developed the linear regression model for LOS on the training data set (based on admissions between Q2 2009 and Q1 2013, n=275,013), and independently confirmed the performance of the resulting model on the validation dataset (based on admissions between Q2 2013 and Q1 2014, n=73,705).

Several changes to predictor variables were made from the previous endorsement. First, the 'PRISM3 Score' is considered to have a non-linear (i.e. quadratic) relationship with the outcome. Previously, PRISM3 Score assumed a piece-wise linear relationship.

Also, the 'age categories' are slightly different (using only neonate (<1m) and infant (1m-1y)) than before (neonate, infant, 1y-3y, 3y-6y, 6y-12y, >12y). The 'age categories' other than neonate and infant where not statistically significant in revised regression model, and found to be 'marginally' statistically significant in the previous regression. While the additional age categories seem reasonable to include, the effect on PICU length of stay for other age groups can be adjusted by other predictors within the model. Therefore, to simplify the model and reduce noise, we removed the additional 'age categories' not statistically significant.

The predictor variable ('Is patient associated with an acute problem?') was added to the 2014 regression model. In the 2014 revision, the variable has a statistical effect in the model ($p<0.0001$) and therefore remained in the regression. In the previous regression model however was found to have a non-statistical effect and removed from the previous model.

Also, one predictor variable ('Did the patient experience head trauma?') was used in the previous regression model, however in 2014 we found it to have a non-statistical effect and removed from the final 2014 model.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of PICU days, PICU days = Number of days between PICU admission and PICU discharge.(For all eligible patients admitted to the ICU, the time at discharge from ICU minus the time of ICU admission (first recorded vital sign on ICU flow sheet)

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Submitted quarterly for all discharges during that time period

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
All patients < 18 years of age

Numerator is the average (mean) observed LOS with the observed LOS (if the observed LOS exceeded 30 days, then the LOS was reduced to 30 days).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is the average (mean) predicted length of stay using the adjustment model.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The denominator is the average (mean) predicted length of stay using the adjustment model.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients => 18 years of age

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patient age > 18 years and patients not eligible for PRISM measurement

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Risk-adjustment measure, not stratification.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Selection criteria for risk adjustment tool for pediatric ICU's:
- Tool must allow quality assessment and comparison between intensive care units, and must be widely used
- Tool must be valid and reliable for severity adjustment and measurement of quality of care provided
- Computation of mortality risk must be in the public domain (i.e. free of charge)
- Algorithms must receive ongoing validation and recalibration

The PRISM 3 model meets these criteria.
VPS has updated the original PRISM LOS model by adding more predictors and re-estimating the coefficients. We developed the linear regression model for LOS on the training dataset (based on admissions between Q2 2009 and

Q1 2013, n=275,013), and independently confirmed the performance of the resulting model on the validation dataset (based on admissions between Q2 2013 and Q1 2014, n=73,705).

A few patients having long ICU stays can disproportionately influence LOS models. We used a 30-day truncation: if any patient had an observed LOS exceeding 30 days, the LOS was reduced to 30 days. Among 348,718 PICU admissions, less than 2% of PICU stays were longer than 30 days.

Since the latest model release is intended to be a refresh of the PRISM III LOS model, we used predictors that are included in PRISM III Risk of Mortality (ROM) and did not include interaction terms or site level predictors. The LOS (in days) is predicted from the following terms at the patient-level:
(1) PRISM3 Score
(2) Neonatal (less than 1 month) patient,
(3) Infant (1 month to 1 year) patient,
(4) Post-operative patient,
(5) Admission of patient from Inpatient Unit,
(6) Previous ICU admission,
(7) Patient with an oncology diagnosis,
(8) Patient with an acute overdose,
(9) Patient with acute diabetes,
(10) Patient with an operative cardiac disease,
(11) Patient with pneumonia,
(12) Patient with non-head trauma,
(13) Patient associated with an acute problem, and
(14) Patient on mechanical ventilation.

References

[1]. Pollack MM. Recalibration of the Length of Stay (LOS) Algorithm: 2006. Personal Communication. 2006.

[2] VPS Webpage. VPS New PRISM 3 LOS Model. 2015.
https://s3.amazonaws.com/vpspublic/PRISM+LOS+brochure.pdf

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
In 2014, the PRISM 3 predicted length of stay model developed by Pollack was recalibrated by VPS. Over 240,000 PICU admissions were used to develop a linear regression model to estimate physical ICU length of stay. ICU length of stay values were truncated at 30 days. The regression model used fifteen predictive variables captured within the first 12 hours of ICU admission, including the PRISM 3 score.[1] The model was extensively validated with an independent dataset of over 81,000 cases. The 2014 predicted length of stay model outperforms the 2006 model.

The standardized length of stay ratio (SLOSR) is created by dividing the sum of observed physical length of stay (truncated at 30 days) by the sum of predicted length of stay. Cases must meet PRISM 3 inclusion criteria to receive a PRISM 3 length of stay prediction.

[1]Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. Critical Care Medicine. 1996 May; 24(5): 743–752.

https://s3.amazonaws.com/vpspublic/PRISM+LOS+brochure.pdf

S.16. **Type of score:**
Ratio
If other:

S.17. **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

S.18. **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 standardized length of stay ratio (SLOSR) is created by dividing the average (mean) observed physical length of stay (truncated at 30 days) by the average (mean) predicted length of stay. Cases must meet PRISM 3 inclusion criteria to receive a PRISM 3 length of stay prediction.

Numerator is the average (mean) observed LOS with the observed LOS = observed LOS exceeding 30 days, the LOS was reduced to 30 days.
The denominator is the average (mean) predicted length of stay using the adjustment model.

Risk adjustment/severity of illness addressed using PRISM 3 methodology.

https://s3.amazonaws.com/vpspublic/PRISM+LOS+brochure.pdf.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. All patients are included.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
No missing data

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*

IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
No mandatory data source or collection instrument for PICU community. Potential resources include PICU-specific databases or the VPS database (myvps.org).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Composite measure, paired measure with readmission (0335)

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0334_MeasureTesting_MSF5.0_12302015..doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0334 NQF Project: Pulmonary Project

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.

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)*: Severity adjusting of LOS is an established method which has been demonstrated to be both valid and reliable, in addition to superior to unadjusted (raw) LOS. Further measure testing is not indicated.

A dataset of 348,718 PICU admissions was split into a training data set (n=275,013) based on admissions

between Q2 2009 and Q1 2013, and an independent dataset for validation (n=73,705) based on admissions between Q2 2013 and Q1 2014.

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

N/A

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

The VPS requires initial and quarterly interrater reliability from all clinical data collectors for each unit participating in VPS. The 2014 aggregate IRR concordance is 96.81%.

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:

Exclusion criteria for the severity adjustment tool (PRISM 3) assures accuracy of severity adjusted LOS calculation.

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)*: Severity adjusting of LOS is an established method which has been demonstrated to be both valid and reliable, in addition to superior to unadjusted (raw) LOS. Further measure testing is not indicated.

An independent dataset of 73,705 PICU admissions between Q2 2013 and Q1 2014 were used to validate the performance of the resulting model.

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

To access the predictive model's validation, the performance was evaluated in four ways using the validation data set.

1. A paired t-test was used to compare the mean observed ICU LOS to the mean predicted ICU LOS for the entire validation population and for various age categories and other subgroups.
2. Coefficients of determination were calculated to measure the variance in LOS by patient and by unit. A regression of the mean observed LOS against the mean predicted LOS was used to assess the proportion of variation across patients (and again for units) which is explained by the model.

3. Calibration curves comparing the average expected LOS to the average observed LOS by dividing the validation data set into 10 equal sized cohorts (i.e. deciles of predicted LOS)
4. Evaluating the model calibration by checking that the distribution of the PICUs' SLOSRs

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)*:

1. The overall t-test shows that the mean difference between the observed LOS and the predicted LOS is not statistically different (6.7 minutes; p=0.79). Similarly for the age categories, no statistical significant differences were found for age the groups '<1 months', '1 months - 23 months', '2 years – 5 years', '6 years – 12 years', '12 years – 18 years'.
2. The predictive model had an R-squared of 0.23 at the patient-level. At the ICU level, the predictive LOS model had an R-squared of 0.92.
3. The calibration curve shows good model fit with R-square of 0.94 (Figure 1).
4. The PICU-level SLOSRs have a normal distribution centered at 1.0 (p=0.15) (Figure 2)

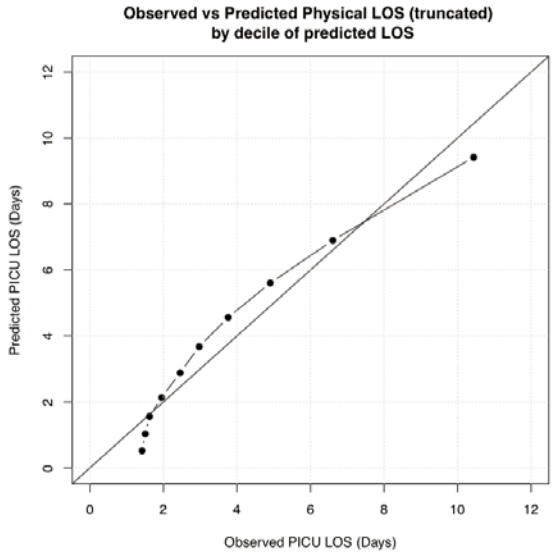


Figure 1: Evaluation of predictive LOS model calibration at the patient level, by deciles of predicted LOS.

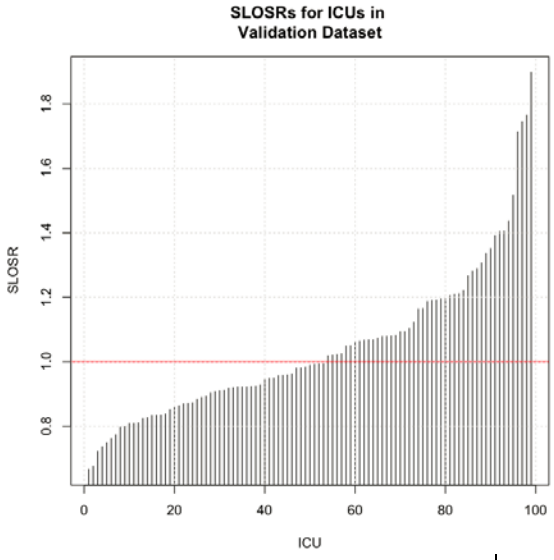


Figure 2: Evaluation of predictive LOS model calibr unit level, by distribution of SLOSRs.

The previous risk adjusted PICU length of stay endorsement predicted the 'natural logarithm of time in ICU'. Therefore an inverse transformation was required to make predictions in real units (days). The 2014 revision used a linear regression modeled directly on the physical PICU length of stay in days –not the log days—so no further transformation is needed to compare observed length of stay with risk-adjusted length of stay.

Seen in Figure 1, the 2014 model is better calibrated than the previous iteration (grouped by predicted LOS deciles). Figure 2 shows better calibration for patient age (grouped by age deciles). This 2014 simplification also allows for easy clinical interpretation (i.e. effects of predictor variables are in days, not log days).

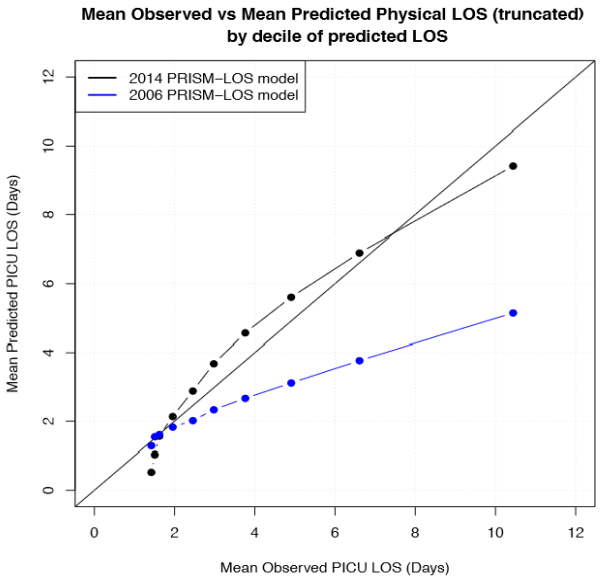


Figure 1 Evaluation of PRISM-LOS model calibration, by deciles of predicted LOS (based on 10 equal sized groups).

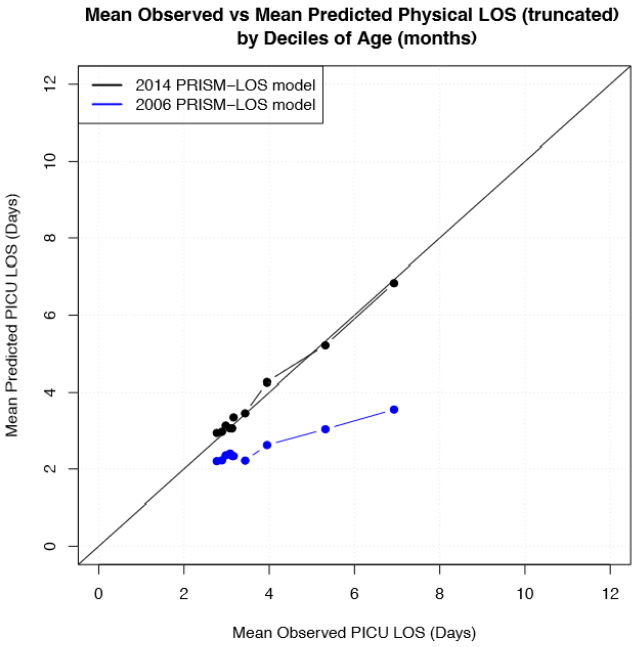


Figure 2 Evaluation of PRISM-LOS model calibration, by deciles of age (based on 10 equal sized groups).

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)*:

The endorsed severity adjustment methodology used for calculating severity adjusted LOS excludes:

• PICU patients >=18 years of age

• PICU patients under the age of 18 years with a stay < 2 hours in the PICU or < 2 consecutive sets of vital signs consistent with life

• Patients admitted to PICU for palliative care

Palliative cases are excluded because the intention is that the patient will likely die in the ICU, skewing SMR calculations if the patients are included.

The other exclusions are consistent with the use of the PRISM 3 instrument for severity adjustment. The tool has not been validated in patients <36 weeks gestation, > or equal to 18 years, or if not in the PICU at least two hours/for two vital signs to be taken.

Further validation of these exclusions is not indicated.

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

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

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)*:

PRISM 3 is a valid, reliable and internationally accepted risk measurement tool. The methodology and measure specifications have been published(1) and are available at https://portal.myvps.org/document/NQFMeasures.pdf

Calibration reassessment has been performed with plans for future model enhancement

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

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)*:

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to

justify lack of adjustment:

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)*:

No sampling was done. However, the data available from the VPS system reveals that the severity adjusted length of stay among 80 participating PICUs ranged from 1.71 to 4.02 days in the third quarter of 2011. This indicates that there is unit specific variance. As numerators, denominators and all definitions are standardized with an IRR >96%, this variation reflects differences in care and not the measurement itself.

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

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

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)*:

Existing literature identifies the shortcomings of not severity adjusting LOS. Thus no attempt to report unadjusted LOS has been made.

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

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)*:

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): N/A. This is consistent with published literature.

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

2.1-2.3 Supplemental Testing Methodology Information:

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. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

All necessary data may be available electronically if an organization has implemented an EHR. In the absence of an EHR, manual data collection would be required.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
The data necessary for capturing length of stay in relatively simple and not burdensome. Data collection for the severity adjustment component is not significant but quite feasible.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees or licensing to collect and calculate this proportion. The VPS provides these data free of charge to participating member units. The 2016 VPS participation fee schedule consists of a baseline fee with incremental charges based on unit volume, thus small units pay $ 16,590 per annum and the largest units pay $33,150 per annum. 2016 VPS fee schedule attached in appendix A1.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Texas Children´s Hospital http://www.texaschildrens.org/About-Us/Quality-Measures/Pediatric-Intensive-Care-Unit/ Payment Program California Children´s Health Services

	http://www.dhcs.ca.gov/services/ccs/Pages/default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) California Children´s Health Services http://www.dhcs.ca.gov/services/ccs/Pages/default.aspx Quality Improvement (Internal to the specific organization) While VPS currently provides data to 128 U.S. PICUs/CICUs from 116 hospitals which use the measure to benchmark against a VPS reference group of other participants. This measure is used for internal Quality Improvement at the individual site level. https://myvps.org

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

While VPS currently provides data to 128 U.S. PICUs/CICUs from 116 hospitals which use the measure to benchmark against a VPS reference group of other participants. This measure is used for internal Quality Improvement at the individual site level. Many of these facilities report patient safety data. An example of one of the participating hospitals listed above list this benchmark data on their hospital-specific web sites.

California Children´s Health Services(CCHS/CCS) includes 28 California based PICUs which participate in VPS and contribute data specific to this measure. In 2014, 14 of the 28 sites contributed data by the California deadline resulting in a state specific performance report to California Children´s Services.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Because of recalibration of the tool there is no comparison data available for measuring improvement. However, this is now being tracked going forward and VPS will be able to address this over next year.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
LOS is critically important particularly when balanced with unplanned readmissions to the ICU. Specifically, length of stay reflects ICU utilization, a source of significant cost and efficiency.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Manual data abstraction with entry into a multi-institutional clinical PICU database (the VPS (myvps.org) has been completed for the variables used in this measure since 2002. Currently, 116 hospitals and 124 PICUs are abstracting and entering data with an aggregate interrater reliability of 96.81%

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** VPS_Fee_Schedule_CY16.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Virtual PICU Systems, LLC
Co.2 **Point of Contact:** Nancy, Brundage, nbrundage@myvps.org, 888-999-4850-103
Co.3 **Measure Developer if different from Measure Steward:** Virtual PICU Systems, LLC
Co.4 **Point of Contact:** Matt, Scanlon, mscanlon@mcw.edu, 888-999-4850-

Additional Information

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

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 12, 2015
Ad.4 **What is your frequency for review/update of this measure?** 3 years
Ad.5 **When is the next scheduled review/update for this measure?** 01, 2016

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** The severity-adjusted length of stay is calculated using a proprietary adjustment tool.

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information
NQF #: 0335 De.2. **Measure Title:** PICU Unplanned Readmission Rate Co.1.1. **Measure Steward:** Virtual PICU Systems, LLC De.3. **Brief Description of Measure:** The total number of patients requiring unscheduled readmission to the ICU within 24 hours of discharge or transfer. 1b.1. **Developer Rationale:** This measure is a critical balancing measure for use with measure 0334 (severity adjusted LOS). Theoretically, units who have lower severity adjusted LOS due to premature discharge of patients would not be identified without pairing the LOS measure with a measure of unplanned readmissions.
S.4. **Numerator Statement:** Total number of unplanned readmissions within 24 hours after discharge/transfer from the PICU. S.7. **Denominator Statement:** 100 PICU Discharges, <18 yrs of age S.10. **Denominator Exclusions:** Patients =>18 years of age,
De.1. **Measure Type:** Outcome S.23. **Data Source:** Electronic Clinical Data : Registry S.26. **Level of Analysis:** Facility
IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Jul 31, 2012
IF this measure is included in a composite, NQF Composite#/title: **IF this measure is paired/grouped, NQF#/title:** De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.
Criteria 1: Importance to Measure and Report
1a. Evidence **Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.**

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This outcome measure indirectly measures process (decision making related to PICU discharge), while directly measuring PICU resource utilization due to unplanned readmission.
- The developer notes this measure should be paired with #0334 (severity-adjusted LOS), indicating it is a critical balancing measure. The developer states that, theoretically, units who have lower severity adjusted LOS due to premature discharge of patients would not be identified without pairing the LOS measure with a measure of unplanned readmissions.

Guidance from the Evidence Algorithm : 1→2 (eligible for PASS rating)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The developer attests the underlying rationale and evidence for this outcome measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

- Current performance data are derived from 82 PICUs submitting data for patients discharged in 2014; 101,295 individual PICU discharges and transfers were recorded. Of these cases, 882 were unscheduled readmissions within 24 hours.
- The unit-level unscheduled readmission rate ranges between 0% and 1.67%. During 2014, the median unit-level unscheduled readmission rate was 0.79%, with interquartile rate of 0.40%-1.09%. The mean unit-level unscheduled readmission rate was 0.78% with standard deviation of 0.42%.
- The patient-level mean unscheduled readmission rate for 2014 was 0.87% (95% CI, 0.81%-0.93%)
- For performance over time: An analysis of 294,204 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e., no increasing or decreasing trend) for unscheduled readmission by quarter (Spearman's correlation test p=0.78).

Disparities:
The developer provides the following:

- Historically population differences have not been found to be variable in pediatric intensive care therapies. One study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status.
    - After adjustment for differences in illness severity, standardized mortality ratios and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care

unit.
    - Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.
- VPS, the developer, stratifies by race/ethnicity, age groups, gender, and insurance payer.
    - Except for two categories, no statistically significant differences were observed for the age categories ('<1 Month', '1 Month - 23 Month', '2 Years - 5 Years', '6 Years - 12 Years', '13 Years - 18 Years'). First, the age group '1 month-23 months' (n=30,130) had a higher unscheduled readmission rate (1.1%; 95% CI, 1.0% - 1.2%, p<0.0001) than all other age groups. Second, the age group '13 years-18 years' (n=23,236) had a lower unscheduled readmission rate (0.67%; 95% CI, 0.57% - 0.78%, p=0.0002) than all other age groups.
    - No statistically significant differences for the rate of unscheduled readmission was found for the following race/ethnic groups: 'African American', 'American Indian/Indigenous', 'Asian/Indian/Pacific Islander', 'Caucasian/European Non-Hispanic', 'Hispanic', and 'Other/Mixed'. Race and ethnicity is reliably collected by 85% of sites.
    - No statistically significant differences for the rate of unscheduled readmission was found for the following insurance statuses: 'Managed Care', 'Commercial/Indemnity Insurance', 'Medicaid & Medicaid Managed Care', 'Self-pay', and 'Other'. Primary payer is reliably collected by 38% of sites.
    - The developer concludes it is not surprising that younger age groups had a higher rate of unplanned readmission in light of their increased vulnerability, as well as physiological differences. The measure itself is not currently stratified by age, but the developer indicates this may provide the ability for enhancement of the measure by age stratification.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The developer indicates that this is an outcomes measure. The evidence is clear and the data elements minimal as it is a PICU readmission within 24 hours of discharge. The outcome is directly related to the data being measured. Changes in practice could potentially decrease readmissions and improve outcomes for patients.

**reasonable evidence, this measure is noted to be a balancing measure for the PICU LOS measure by same company.

My concerns with this are similar for all VPS metrics. They are proprietary and SOI adjustment is not based on most recent evidence (less relevant for this measure but rather the other two VPS measures)

**#0335 measures the results of a process (efficient and effective PICU discharge) rather than a patient outcome (unless we count PICU readmission as a patient outcome). The evidence implies that a PICU readmission rate is inversely proportional to discharge efficiency and a high readmission rate may be the result of a faulty PICU discharge process (timing of discharge, severity threshold level for discharge, and/or patient selection). Although this is a common sense conclusion, I am unconvinced that it is fully supported by the evidence.

1b. Performance Gap

Comments:
**There was limited evidence provided that identified a gap in care based on population (race, ethnicity, urban or rural, etc.). Even with evidence of race disparities, the developer chose not to utilize the information in the measure development. However, the results of the measure clearly identify an opportunity exists for improvement to achieve optimal performance.
**If used as a balancing measure, this metric may provide value. AS a stand along measure this is less valuable for cross hospital comparisons. Many unit based factors likely lead to PICU readmissions. perhaps this and the LOS measure should be combined.
**Yes. Relatively small overall readmission rate given the remarkably broad range of disease processes routinely seen in the PICUs, variability of severity, and heterogenous availability of subspecialists (as one example variable) between hospitals and units. Rates were remarkably similar with a narrow range, so not sure that this warrants a national measure. Regarding disparities, the only real differential was between "uninsured" children, who appeared to be sicker upon admission and had shorter PICU stays (increased mortality?), and those "insured." This is a "pre-admission bias and not really controllable by the units. Not sure disparities warrant a national measure either.

1c. High Priority (previously referred to as High Impact)
Comments:
**The measure is not a composite performance measure.
**na
**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data : Registry

Specifications:

- The developer attests the measure specifications have not been updated since the last review.
- The measure tracks the total number of unplanned readmissions within 24 hours after discharge/transfer from the PICU per 100 PICU <18 years of age discharges.
- The calculation algorithm is stated in S.18 and appears straightforward.

Questions for the Committee :

- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The previous submission noted the following: As this measure is a simple proportion using previously established methods, there is no further reliability or validity assessment that is indicated.

- The developer states elsewhere, however, that "numerators, denominators and all definitions are standardized with an inter-rater reliability (IRR) >96%." From this one could infer validity testing at the data element level was assessed. Per NQF guidance, separate reliability testing is not required when validity testing at the data element level is performed for all critical data elements.

Describe any updates to testing:

- The developer attests no new testing was performed for this submission.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Results of reliability testing:

- The results of the data element level validity testing are provided in the following section.

Guidance from the Reliability Algorithm : Not applicable

Question for the Committee:

- *The developer attests that the specifications have not changed and new testing has not been performed since the last NQF endorsement review. Does the Committee believe there is no need for repeat discussion and vote on Reliability?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The previous submission noted the following: As this measure is a simple proportion using previously established methods, there is no further reliability or validity assessment that is indicated.
- The developer states elsewhere, however, that "numerators, denominators and all definitions are standardized with an inter-rater reliability (IRR) >96%." From this one could infer validity testing at the data element level was assessed.
    - The developer does not provide any additional information on the reliability of specific

data elements.
    - Only reference to inter-rater reliability is made. The developer does not provide additional analyses (e.g., sensitivity, specificity, PPV, NPV).

Describe any updates to validity testing

- The developer attests no new testing was performed for this submission.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer describes its IRR methodology, as follows: "Inter-rater reliability (IRR) is a certification method to assure standardization in data collection. Data collectors collect the VPS data fields from the patient medical record in accordance with the VPS definitions and independently from the remainder of their team. Results are compared amongst the team members for areas of concordance. A 90% concordance must be obtained for a team to pass the IRR and become certified to collect data. The process is done initially upon joining the VPS and quarterly thereafter to maintain certification."

Validity testing results:

- The developer states, "numerators, denominators and all definitions are standardized with an IRR >96%."

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *The developer attests that the specifications have not changed and new testing has not been performed since the last NQF endorsement review. Does the Committee believe there is no need for repeat discussion and vote on Validity?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Exclusions include:
    - Patients = 18 years of age
    - Readmissions > 24 hours following discharge/transfer from PICU
    - All planned readmissions
        - Planned readmissions are excluded as they reflect conscious care decisions rather than potential failures of care processes.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☒ **Statistical model** ☐ **Stratification**

- There is no risk adjustment or stratification.
- The developer provides the following justification for the lack of risk adjustment: "This is simply a proportion of readmissions. While secondary analysis to see if these were due to risk factors including severity of illness, by virtue of the perception that they were ready for discharge, severity of illness at time of discharge is not routinely calculated nor would it be a valid process."

Questions for the Committee:

- *Does the Committee concur with the developer's rationale for no risk adjustment or stratification.*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer provides the following:

- Current data are derived from 82 PICUs submitting data for patients discharged in 2014; 101,295 individual PICU discharges and transfers were recorded. Of those cases, 882 were unscheduled readmissions within 24 hours.
- The unit-level unscheduled readmission rate ranges between 0% and 1.67%. During 2014, the median unit-level unscheduled readmission rate was 0.79% with interquartile rate of 0.40% - 1.09%. The mean unit-level unscheduled readmission rate was 0.78% with standard deviation of 0.42%.
- The patient-level mean unscheduled readmission rate for 2014 was 0.87% (95% CI, 0.81%-0.93%)
- For performance over time: An analysis of 294,204 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e. no increasing or decreasing trend) for unscheduled readmission by quarter (Spearman's correlation test p=0.78).

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- Missing data analyses were not conducted for this measure

Guidance from the Validity Algorithm: 1→2 →3→6→10→11 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Data validity testing was conducted at the data element level through inter-rater reliability testing. The specifications are consistent with the evidence.

**reasonable

**None

2a2. Reliability Testing

Comments:

**• The developer describes its IRR methodology, as follows: "Inter-rater reliability (IRR) is a certification method to assure standardization in data collection. Data collectors collect the VPS data fields from the patient

medical record in accordance with the VPS definitions and independently from the remainder of their team. Results are compared amongst the team members for areas of concordance. A 90% concordance must be obtained for a team to pass the IRR and become certified to collect data. The process is done initially upon joining the VPS and quarterly thereafter to maintain certification."
**Similar comments as above
**The number of entities and patients was likely sufficient for broad generalization but I remain concerned that specific decision-making elements (leading to successful and un-successful PICU discharges) were not teased out. The assumption here is that mistakes made deciding who may and may not be successfully discharged from the PICU directly relate to quality of care. This assumption is intuitively valid but I remain unconvinced that these results demonstrate sufficient validity so that conclusions about quality can be made. Again, intuitively the score from this measure as specified is an indicator of quality but there are also variables (e.g., quality of post-PICU care, etc.) which directly impact the numerator which might not reflect the QOC in the PICU or w.r.t. the original discharge decision.

2b2. Validity Testing
Comments:
**Administrative data is used for this measure calculation. Missing data/no response is not considered to be a concern.
**Planned definition application has potential to vary across sites
**The exclusions (>= 18 yo, readmissions >24h p discharge/transfer from PICU, and planned readmissions) are not arbitrary and are definitional. I am concerned that there is no risk adjustment or leveling - it presumes that a PICU is a PICU, when in reality each hospital and PICU has its own "character" and degree and type of specialization (= different patient populations). Without a way to "normalize" admissions, does this prevent us from coming up with clinically meaningful differences in the re-admissions rates between PICUs?

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**No changes were made to the measure that required reliability testing. Since the validity was tested at the data element level, NQF does not require additional reliability testing for this measure.
**What is the criteria for updating this information? It would seem to be a reasonable expectation for these to be retested. Specifically determining which admissions planned or unplanned seem possible to vary across sites.
**My concern here is that the overall readmission rate is so low that even an IRR "unreliability rate" of <=4% could have a statistical impact. Conducted at data element level - not sure what effect this might have had at a scoring level.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer provides the following information:

- Some data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The availability of data in electronic sources depends on each individual organization´s resources. There is no reason the data could not be readily available in an implemented EHR.
- There are no fees or licensing required to collect and calculate this proportion. There are potential

fees associated with using available software products such as the VPS; the 2016 fee schedule is attached.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments:

**The measure is calculated using administrative data. It could potentially be calculated as an e measure as this capability improves.
**seems feasible
**The required data element of "unplanned readmission" is not currently routinely generated and used during care delivery. A post-facto decision must be objectively made. This required data element is not available in electronic form. It concerns me that the data collection strategy (chart review by objective reviewer(s)) may be difficult to put into practical operation.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- The measure data are not aggregated and publicly reported; however, some hospitals participating in the VPS system (e.g., in Cleveland and Wisconsin) may individually publicly report their data.
- The developer stated that the funding body for California pediatric healthcare, California Children's Services, mandates use for accountability purposes.
- VPS currently provides data to 128 U.S. PICUs/CICUs from 116 hospitals, which use the measure to benchmark against a VPS reference group of other participants.

Improvement results:

- From 1/1/2012 until 12/31/2014 the quarterly unscheduled readmission rate for VPS's U.S. PICUs (PICUs only) has varied between 0.68% - 0.96% (n=294,204). The developer state there is no increasing or decreasing trend for the overall rate of unscheduled readmission (Spearman

correlation test, p=0.78).

Unexpected findings (positive or negative) during implementation:

- The developer states there were no unexpected findings during implementation.

Potential harms:

- The developer states that, theoretically, there is a risk of miscapturing the time of original discharge and thus incorrectly including or excluding a patient. Additionally, there is a risk for potential error related to misidentifying readmissions as either planned or unplanned.

Feedback: No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Is use only in local programs sufficient? Should the measure be publicly reported on a broader scale?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**It is of concern that very little improvement is being made through use of this measure. This could be related to the interventions being applied. The measure data are not aggregated and publicly reported; however, some hospitals participating in the VPS system (e.g., in Cleveland and Wisconsin) may individually publicly report their data.

The developer stated that the funding body for California pediatric healthcare, California Children's Services, mandates use for accountability purposes. VPS currently provides data to 128 U.S. PICUs/CICUs from 116 hospitals, which use the measure to benchmark against a VPS reference group of other participants.

**Similar comments in this section to other PICU Measures we are considering. Is a couple hospitals posting on their own websites adequate for stating that there is "public reporting?"

Does the fact that this is not a trend in decreasing or increasing rates impact the utility of this measure?

**Realistically, the measure is only being publicized voluntarily; presumptively by those who "score well." In two years of data collection, there were no statistically appreciative changes in the readmission rates, so the measure does not seem to be affecting changes. But the measure also does not point toward what, specifically, would need to be changed by the provider in order to affect positive quality-of-care change. Not sure just how this measure would be helpful...

Criterion 5: Related and Competing Measures

Related or competing measures

- 0334: PICU Severity-adjusted Length of Stay

Harmonization

- Developer indicates this measure and 0334 have been paired. The measure specifications are

harmonized with NQF 0334.

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0335 NQF Project: Pulmonary Project

1. IMPACT, OPPORTUITY, 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)

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 indirectly measures process (decision making related to PICU discharge) while directly measuring PICU resource utilization due to unplanned readmission.

This measure indirectly measures process (decision making related to PICU discharge) while directly measuring PICU resource utilization due to unplanned readmission. CMS has identified the importance of 30 day hospital admission (see below). However, the important CMS measure maybe insensitive to ICU management decisions. Therefore the 24 hour PICU readmission is complementary to the CMS measure.

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

Selected individual studies (rather than entire body of evidence)

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)*:

A number of studies have specifically measured the rate and impact of unplanned ICU admissions. Literature sites mortality is higher in readmission cases.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: Six studies since 2005 have specifically looked at early, unplanned readmission to the ICU setting.

Multiple studies since 2005 have specifically looked at early, unplanned readmission to the ICU setting.

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):* The studies look at different populations making pooling of findings problematic. The time frame to be included as a readmission varies in the studies.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* While the studies varied in the inclusion criteria, the findings all indicated increased risk of morbidity and mortality among this cohort

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

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:

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

1c.12 If other, identify and describe the grading scale with definitions: per 1c.9 above no grading has been done.

1c.13 Grade Assigned to the Body of Evidence: N/A

1c.14 Summary of Controversy/Contradictory Evidence: None

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

Brunetti MA, Glatz AC, McCardle K, et al. Unplanned Readmission to the Pediatric Cardiac Intensive Care Unit: Prevalence Outcomes, and Risk Factors. World J Pediatr Congenit Heart Surg. 2015 Oct;6(4):597-603.

Cunha F, Teixeira-Pinto, A. Back to the PICU: Who Is at Risk and Outcome of Unplanned Readmissions. Critical Care Medicine 2013; 41(12) 2831-2832.

Edwards J, Lucas A, Stone P, et al. Frequency, Risk Factors, and Outcomes of Early Unplanned Readmissions to PICUs. Critical care medicine. 2013;41(12):2773-2783. doi:10.1097/CCM.0b013e31829eb970.

Kramer AA, Higgins TL Zimmerman JE. Intensive care unit readmissions in U.S. hospitals: patient characteristics, risk factors and outcomes. Crit Care Med 2012; 40(1) 3-10

Haller G, Myles PS, Wolfe R, et al. Validity of unplanned readmission to an intensive care unit as a measure of patient safety in surgical patients. Anesthesiology. 2005; 103(6) 1121-9

Makris N, Dulhunty JM Paratz JD et al. Unplanned early readmission to the intensive care unit: a case-control study of patient, intensive care and ward-related factors. Anaest Intesive Care; 2010 38(4): 732-31

Mandell I, Bynum F, Marshall L, et al. Pediatric Early Warning Score and unplanned readmission to the pediatric intensive care unit. Journal of Critical Care; 2015 30(5): http://dx.doi.org/10.1016/j.jcrc.2015.06.019.

White D, Coscrato Bachmayr H. 808 Unplanned Readmission to the Paediatric Intensice Care Unit (PICU). Can it be Prevented? Arch Dis Child; 2012 97(2)A232: doi:10.1136/archdischild-2012-302724.080.

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

Readmission Measures: Centers for Medicare & Medicaid Services Readmission Reduction Program.
Section 3025 of the Affordable Care Act added section 1886(q) to the Social Security Act establishing the Hospital Readmissions Reduction Program, which requires CMS to reduce payments to IPPS hospitals with excess readmissions, effective for discharges beginning on October 1, 2012. The regulations that implement this provision are in subpart I of 42 CFR part 412 (§412.150 through §412.154).

"In the FY 2012 IPPS final rule, CMS finalized the following policies with regard to the readmission measures under the Hospital Readmissions Reduction Program:

- Defined readmission as an admission to a subsection (d) hospital within 30 days of a discharge from the same or another subsection (d) hospital;
- Adopted readmission measures for the applicable conditions of acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN);
- Established a methodology to calculate the excess readmission ratio for each applicable condition, which is used, in part, to calculate the readmission payment adjustment. A hospital's excess readmission ratio is a measure of a hospital's readmission performance compared to the national average for the hospital's set of patients with that applicable condition.
- Established a policy of using the risk adjustment methodology endorsed by the National Quality Forum

(NQF) for the readmissions measures to calculate the excess readmission ratios, which includes adjustment for factors that are clinically relevant including certain patient demographic characteristics, comorbidities, and patient frailty.

- Established an applicable period of three years of discharge data and the use of a minimum of 25 cases to calculate a hospital's excess readmission ratio for each applicable condition.

In the FY 2014 IPPS final rule, CMS adopted the application of an algorithm to account for planned readmissions to the readmissions measures. In addition, CMS finalized the expansion of the applicable conditions beginning with the FY 2015 program to include: (1) patients admitted for an acute exacerbation of chronic obstructive pulmonary disease (COPD); and (2) patients admitted for elective total hip arthroplasty (THA) and total knee arthroplasty (TKA).

In the FY 2015 IPPS final rule, CMS finalized the expansion of the applicable conditions beginning with the FY2017 program to include patients admitted for coronary artery bypass graft (CABG) surgery in the calculation of a hospital's readmission payment adjustment factor.

In the FY 2016 IPPS final rule, CMS finalized an update to the pneumonia readmission measure by expanding the measure cohort to include additional pneumonia diagnoses: (i) patients with aspiration pneumonia; and (ii) sepsis patients coded with pneumonia present on admission (but not including severe sepsis)".

1c.17 Clinical Practice Guideline Citation: Centers for Medicare & Medicaid Services Readmission Reduction Program.
Section 3025 of the Affordable Care Act. 1886(q) to the Social Security Act. Subpart I of 42 CFR part 412 (§412.150 through §412.154).

1c.18 National Guideline Clearinghouse or other URL: none https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

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: N/A

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

1c.22 If other, identify and describe the grading scale with definitions: per 1c.19 above no grading has been done.

1c.23 Grade Assigned to the Recommendation:

1c.24 Rationale for Using this Guideline Over Others:

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: Low 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0335_Evidence_MSF5.0_data.123115..doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure is a critical balancing measure for use with measure 0334 (severity adjusted LOS). Theoretically, units who have lower severity adjusted LOS due to premature discharge of patients would not be identified without pairing the LOS measure with a measure of unplanned readmissions.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Current performance:
The data from 82 PICUs submitting data for patients discharged in 2014. In total 101,295 individual PICU discharges and transfers were recorded; of those cases, 882 were unscheduled readmissions within 24 hours.
The unit-level unscheduled readmission rate ranges between 0% and 1.67%. Over 2014, the median unit-level unscheduled readmission rate was 0.79% with interquartile rate of 0.40% - 1.09%. The mean unit-level unscheduled readmission rate was 0.78% with standard deviation of 0.42%.
The patient-level mean unscheduled readmission rate for 2014 was 0.87% (95% CI, 0.81%-0.93%)

Performance overtime:
An analysis of 294,204 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e. no increasing or decreasing trend) for unscheduled readmission by quarter (Spearman's correlation test p=0.78).

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
The lack of evidence for improvement does not invalidate the importance of this measure. Instead, this measure

serves as a balancing measure for severity adjusted ICU length of stay. The lack of trend in unplanned readmissions provides reassurance that the severity adjusted ICU length of stay does not reflect gaming of the measure by institutions.

1b .4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Historically population differences have not been found to be variable in pediatric intensive care therapies. A study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status. After adjustment for differences in illness severity, standardized mortality ratios and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care unit. Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.

http://pediatrics.aappublications.org/content/127/3/e588.short
http://link.springer.com/article/10.1007/s10900-014-9823-0#/page-1

We stratify by race/ethnicity, age groups, gender, and insurance payer.

Except for two categories, no statistically significant differences were observed for the age categories ('<1 Month', '1 Month - 23 Month', '2 Years - 5 Years', '6 Years - 12 Years', '13 Years - 18 Years'). First, the age group '1 month-23 months' (n=30,130) had a higher unscheduled readmission rate (1.1%; 95% CI, 1.0% - 1.2%, p<0.0001) than all other age groups. Second, the age group '13 years-18 years' (n=23,236) had a lower unscheduled readmission rate (0.67%; 95% CI, 0.57% - 0.78%, p=0.0002) than all other age groups.

No statistically significant differences for the rate of unscheduled readmission was found for the following race/ethnic groups: 'African American', 'American Indian/Indigenous', 'Asian/Indian/Pacific Islander', 'Caucasian/European Non-Hispanic', 'Hispanic', and 'Other/Mixed'. Race and ethnicity is reliably collected by 85% of sites.

No statistically significant differences for the rate of unscheduled readmission was found for the following insurance statuses: 'Managed Care', 'Commercial/Indemnity Insurance', 'Medicaid & Medicaid Managed Care', 'Self-pay', and 'Other'. Primary payer is reliably collected by 38% of sites.

The fact that younger age groups had a higher rate of unplanned readmission is not surprising in light of their increased vulnerability as well as physiological differences. The measure itself is not currently stratified by age. This may provide the ability for enhancement of the measure by age stratification.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Limitations to evaluations of disparities are common in the research. Literature has shown variations in data by factors associated with differing population groups.

Epstein D, Reibel M, Unger J, et al. The Effect of Neighborhood and Individual Characteristics on Pediatric Critical Illness. Journal of Community Health 2014; 39(4). 753-759.

Epstein D, Wong C, Khemani R, et al. Race/Ethnicity is Not Associated with Mortality in the PICU. Pediatrics 2011; 127(3). e588-e597. http://pediatrics.aappublications.org/content/127/3/e588.short.

Lopez A, Tilford J, Anand K, et al. Variation in pediatric intensive care therapies and outcomes by race, gender and insurance status. Ped Crit Care Med 2006; 7(1). 2-6.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
"The average ICU readmission rate of 7% (range, 4 to 14%) has remained relatively unchanged in both North America and Europe. Respiratory and cardiac conditions were the most common (30 to 70%) precipitating cause of ICU readmission. Patients readmitted to ICUs had average hospital stays at least twice as long as nonreadmitted patients. Hospital death rates were 2- to 10-times higher for readmitted patients than for those who survived an ICU admission and were never readmitted. Predictors of ICU readmission have been neither well studied nor reproducible."(1)
Patient in intensive care units (ICUs) account for nearly 30% of acute care hospital costs, yet these patients occupy only 10% of inpatient beds. Care provided in intensive care units accounts for a large percentage of acute care hospital costs.(2)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Rosenberg AL, Watts C. Patients Readmitted to ICUs-A Systematic Review of Risk Factors and Outcomes. Chest: 2000; 118:492-502.
2. Brilli RJ, Spvetz A, Branson, RD, et al. Critical care delivery in the intensive care unit: Defining clinical roles and the best practice model. Critical Care Medicine; 2001: 29 (10), 2007-2019.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
n/a

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety : Readmissions

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://s3.amazonaws.com/vpspublic/NQFMeasures.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
none, there has been no change in the risk model over time

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Total number of unplanned readmissions within 24 hours after discharge/transfer from the PICU.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Unplanned readmission within 24 hours of discharge/transfer.

Data submission quarterly with reporting on annual basis.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Inclusion: All PICU patients < 18 years of age
Exclusions:
- Patients = 18 years of age
- Readmissions > 24 hours following discharge/transfer from PICU
- All planned readmissions

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
100 PICU Discharges, <18 yrs of age

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
All PICU patients <18 years of age

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients =>18 years of age,

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patients not yet discharged from PICU

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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.)*
First, identify all discharges/transfers from PICU who are readmitted, limited to children <18 years of age.

Second, exclude all planned readmissions.
Third, use above number as numerator over denominator of PICU dischages/transfers.
Report per 100 PICU discharges

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
No sampling used. All discharges meeting inclusion criteria as included in measure.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
none

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
No mandatory data source or collection instrument for PICU community. Potential resources include PICU-specific databases or the VPS database (myvps.org).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0335_MeasureTesting_MSF5.0_Data..01142016..doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0335 NQF Project: Pulmonary Project

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.

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)*:

As this measure is a simple proportion using previously established methods, there is no further reliability or validity assessment that is indicated.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*: N/A

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

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:

As this measure is a simple proportion using previously established methods, there is no further reliability or validity assessment that is indicated.

This specific measure appropriately targets children in PICU settings who are readmitted in an unplanned manner within 24 hours of discharge/transfer requiring an upgrade in the level of care.

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)*: No sampling was done. This indicates that there is unit specific variance. As numerators, denominators and all definitions are standardized with an IRR >96%, this variation reflects differences in care and not the measurement itself. Inter-rater reliability (IRR) is a certification method to assure standardization in data collection. Data collectors collect the VPS data fields from the patient medical record in accordance with the VPS definitions and independently from the remainder of their team. Results are compared amongst the team members for areas of concordance. A 90% concordance must be obtained for a team to pass the IRR and become certified to collect data. The process is done initially upon joining the VPS and quarterly thereafter to maintain certification.

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

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)*: N/A

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)*:

Exclusions include:

- Patients = 18 years of age
- Readmissions > 24 hours following discharge/transfer from PICU
- All planned readmissions

This is appropriate as the measure is intended to describe care of children (<18 years of age) in a PICU setting.

Planned readmissions are excluded as they reflect conscious care decisions rather than potential failures of care processes.

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

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

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)*:

Not applicable (see 2b4.4)

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

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)*: N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: This is simply a proportion of readmissions. While secondary analysis to see if these were due to risk factors including severity of illness, by virtue of the perception that they were ready for discharge, severity of illness at time of discharge is not routinely calculated nor would it be a valid process.

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)*: Unscheduled readmission rate

within 24 hours for data submitted by 80 PICUs for Q3 2011 are between 0% and 3.14%.

Current performance:

The data from 82 PICUs submitting data for patients discharged in 2014. In total 101,295 individual PICU discharges and transfers were recorded; of those cases, 882 were unscheduled readmissions within 24 hours.

The unit-level unscheduled readmission rate ranges between 0% and 1.67%. Over 2014, the median unit-level unscheduled readmission rate was 0.79% with interquartile rate of 0.40% - 1.09%. The mean unit-level unscheduled readmission rate was 0.78% with standard deviation of 0.42%.

The patient-level mean unscheduled readmission rate for 2014 was 0.87% (95% CI, 0.81%-0.93%)

Performance overtime:

An analysis of 294,204 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e. no increasing or decreasing trend) for unscheduled readmission by quarter (Spearman's correlation test p=0.78).

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*: Site performance is identified by test for equality of proportions of patients readmitted.
2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: See 2b5.1

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)*: No sampling was done. However, the data available from the VPS system reveals that the unplanned readmission rates among 80 participating PICUs ranged from 0% to 3.14% in the third quarter of 2011. This indicates that there is unit specific variance.

No sampling was done

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

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)*: N/A

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)*: N/A. This is consistent with published literature.

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

2.1-2.3 Supplemental Testing Methodology Information: N/A

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. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other
If other: capture from electronic sources (eg, time of discharge orders)

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**
The availability of data in electronic sources depends on each individual organization´s resources. There is no reason the data could not be readily available in an implemented EHR.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
The identification of unplanned readmissions within 24 hours is a straightforward process that requires minimal resources.

The current VPS cohort indicates that during the time of 1/1/2009-6/30/2015, VPS has 598,598 unique pediatric ICU admissions (<18 years old), from 145 PICUs (includes 12 CICUs) located in the US. Individual sites manage the data collection process within their own institutions and submit data to VPS quarterly.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees or licensing to collect and calculate this proportion. There are potential fees associated with using available software products such as the VPS (see Appendix A1).

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Children´s Hospital Wisconsin http://www.chw.org/ Cleveland Clinic Children´s Hospital http://my.clevelandclinic.org/ccf/media/files/outcomes/2013/outcomes-peds.pdf Texas Children´s Hospital http://www.texaschildrens.org/About-Us/Quality-Measures/Pediatric-Intensive-Care-Unit/ Payment Program California Children´s Health Services http://www.dhcs.ca.gov/services/ccs/Pages/default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) California Children´s Services http://www.dhcs.ca.gov/services/ccs/Pages/default.aspx Quality Improvement (Internal to the specific organization) 116 pediatric ICUs currently receive VPS data as part of their internal QI evaluation. http://myvps.org

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

VPS currently provides data to 128 U.S. PICUs/CICUs from 116 hospitals which use the measure to benchmark against a VPS reference group of other participants. This measure is used for internal Quality Improvement at the individual site level. Many of these facilities report patient safety data. An example of three of the participating hospitals listed above list this benchmark data on their hospital-specific web sites.

California Children´s Health Services(CCHS/CCS) includes 28 California based PICUs which participate in VPS and contribute data specific to this measure. In 2014, 14 of the 28 sites contributed data by the California deadline resulting in a state specific performance report to California Children´s Services.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is

demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

From 1/1/2012 until 12/31/2014 the quarterly unscheduled readmission rate for VPS's U.S. PICUs (PICUs only) has varied between 0.68% - 0.96% (n=294,204). No increasing or decreasing trend for the overall rate of unscheduled readmission (Spearman correlation test, p=0.78).

Geographic area includes: 44 states, 145 PICU/CICUs

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Previously, analysis of data from 80 PICUs submitting data in Q3 2011 to the VPS system revealed unplanned readmission rates ranging from 0% to 3.14% of discharged patients.

Current analysis displays data from 82 PICUs submitting data from 2014 to the VPS system revealed unplanned readmission rates range from 0 % to 1.67%.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Theoretically, there is a risk of miscapturing the time of original discharge and this incorrectly including or excluding a patient. Additionally, there is a risk for potential error related to misidentifying readmissions as either planned or unplanned.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)
All existing and potentially competing measures endorsed by NQF are 1) focused on adults and 2) focused on hospital populations with an emphasis on readmission to the hospital, not the ICU. They are fundamentally different in their intent.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** VPS_Fee_Schedule_CY16-635883941256952289.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Virtual PICU Systems, LLC
Co.2 **Point of Contact:** Nancy, Brundage, nbrundage@myvps.org, 888-999-4850-103
Co.3 **Measure Developer if different from Measure Steward:** Virtual PICU Systems, LLC
Co.4 **Point of Contact:** Matt, Scanlon, mscanlon@mcw.edu, 888-999-4850-

Additional Information

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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2008

Ad.3 **Month and Year of most recent revision:** 12, 2015

Ad.4 **What is your frequency for review/update of this measure?** 3 years

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information
NQF #: 0343 De.2. **Measure Title:** PICU Standardized Mortality Ratio Co.1.1. **Measure Steward:** Virtual PICU Systems, LLC De.3. **Brief Description of Measure:** The ratio of actual deaths over predicted deaths for PICU patients. 1b.1. **Developer Rationale:** Use of SMR allows for comparing care in different PICUs while allowing for accounting for variations in the severity of patients´ illness. While there is no "right" mortality rate for a PICU, comparisons of SMR can guide quality improvement efforts.
S.4. **Numerator Statement:** Actual number of deaths occurring in PICU. S.7. **Denominator Statement:** The sum of of predicted PRISM 3 mortality. "Predicted mortality" = Number of deaths expected based on assessed physiologic risk of mortality. Include all PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment or boarder/IMCU status. S.10. **Denominator Exclusions:** Include all PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment or boarder/IMCU status.
De.1. **Measure Type:** Outcome S.23. **Data Source:** Administrative claims, Electronic Clinical Data : Registry, Paper Medical Records S.26. **Level of Analysis:** Facility
IF Endorsement Maintenance – Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Jul 31, 2012
IF this measure is included in a composite, NQF Composite#/title: **IF this measure is paired/grouped, NQF#/title:** De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This measure relates to a health outcome, mortality.
- The standardized mortality ratio is simply a ratio. It is accepted as an appropriate quality measure for ICU settings. The developer reports the literature is consistent on the value of using SMR with three caveats.
    - Use of a calibrated tool for severity adjustment has been identified as important.
    - A recent publication by Brinkman in Critical Care Medicine (2012; 40(2)373-378) identified that use of a physiology-based tool when calculating SMR is superior than using administrative data.
    - Premature transfer of patients from an ICU can lower the SMR, creating a potential for "gaming" measures (Kahn Chest. 2007;131(1):68-75). Use of measure #343 in conjunction with measure #0334 and #0335 addresses the potential for gaming.
- The previous Committee concluded the literature demonstrates that a standardized mortality ratio is an appropriate measure for ICU settings.
- The developer attests the evidence has not changed since the last endorsement maintenance review.

Guidance from the Evidence Algorithm : 1→2 (eligible for PASS rating)

Questions for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The developer attests the evidence and underlying rationale for this outcome measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

- Current performance data are derived from 79 PICUs submitting PRISM III data for patients discharged in 2014; 92,943 individual PICU discharges and transfers were recorded.
- The unit-level standardized mortality ratio (SMR) ranges between 0.16 and 2.02. During2014, the median unit-level SMR was 0.92 with interquartile rate of 0.79-1.15. The mean unit-level SMR was 0.97, with standard deviation of 0.32.
- The patient-level mean SMR for 2014 is not statistically different from 1.0 (0.99; 95% CI, 0.96 – 1.02).

- For performance over time: An analysis of 273,130 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e. no increasing or decreasing trend) for SMR by quarter (Spearman's correlation test p=0.39).

Disparities

The developer provides the following:

- Population differences have not been found to be variable in pediatric intensive care therapies. One study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status.
    - After adjustment for differences in illness severity, SMRs and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care unit.
    - Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.
- VPS stratifies by race/ethnicity, age groups, gender, and insurance payer.
    - Except for two categories, no statistically significant differences for SMR were observed for the age categories ('<1 Month', '1 Month - 23 Month', '2 Years - 5 Years', '6 Years - 12 Years', '13 Years - 18 Years'). First, the age group '1 month-23 months' (n=27,732) had a higher SMR (1.08; 95% CI, 1.02 - 1.14, p=0.0002) than all other age groups. Second, the age group '13 years-18 years' (n=21,445) had a lower SMR (0.89; 95% CI, 0.82 - 0.96, p=0.0021) than all other age groups.
    - No statistically significant differences for the SMR was found for the following race/ethnic groups: 'African American', 'American Indian/Indigenous', 'Asian/Indian/Pacific Islander', 'Caucasian/European Non-Hispanic', 'Hispanic', and 'Other/Mixed'. Race and ethnicity is reliably collected by 85% of sites.
    - No statistically significant differences for the SMR was found for the following insurance statuses: 'Managed Care', 'Commercial/Indemnity Insurance', Medicaid & Medicaid Managed Care', 'Self-pay', and 'Other'. Primary payer is reliably collected by 38% of sites.
    - No statistically significant differences for the SMR were found between gender/sex.
    - The developer states the higher SMR in lower age groups may reflect physiologic differences or vulnerability of this population related to their dependency on care givers pre-hospitalization. Similarly the lower SMR in the teen-aged population may reflect physiologic differences.

Question for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The measure is stated to be an outcomes measure. Essentially it compares mortality rates in a PICU with a goal of reducing the rate by measuring outcomes. Outcome rates do not indicate that measuring has resulted in any change in outcomes.

**The evidence suggests good relationship to the outcome. This is a reasonable outcome measure, albeit one of low frequency. Agree that the evidence has not tangibly changed since last submission. Only change is an update to the severity of illness assignment algorithm, however it should not impact this current measure.

**#0343 relates tangentially to the process of caring for critically ill children in PICUs - theoretically, QOC is inversely proportional to the mortality rate divided by the predicted/expected mortality rate based on a commercially available scoring system. There is no identified relationship between this measured outcome and any specific health care action(s). I fail to see a direct cause and effect. While the hypothesis is a common sense assumption, the literary evidence cannot prove causation within a specific PICU or on a specific case.

1b. Performance Gap

Comments:

**The developer did not indicate any performance gap based on population subgroups. PICU's were staffed based on intensity of care needs as determined by the evidence.

**Performance data provided and shows variability in care. The developers note no differences in population subtypes other than age. As they point out, those age differences likely reflect physiological differences more than ICU performance differences.

**Performance data was provided over a 3-year period, to the end of 2014. The unit-level SMR suggests a gap between the PICUs with few deaths at all versus those units with twice the expected number. This may not be reflective of the "character" of the unit and types of patients and referral patterns. The mean was just where it would be expected to be. Once "corrected" for the severity of the patient population, though, this washes out and the units equalize - i.e., no gap. Interestingly, over 3 years of transparency, there was still no trending of the data. It is difficult to see then, based on this information, how this warrants a national measure. The demographics do not suggest care disparities with the exception of the pre-care insurance status. Kids without insurance were sicker when they presented, had shorter PICU stays, and higher mortality, across the board.

1c. High Priority (previously referred to as High Impact)

Comments:

**This is not a composite measure.

**na

**n/a

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data : Registry

Specifications:

- The developer attests the measure specifications have not been updated since the last review.
- The measure tracks the actual number of deaths occurring in PICU and includes all PICU patients <18 years of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment or boarder/IMCU status.
- No calculation algorithm is stated in S.18.

Questions for the Committee :

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing **Testing attachment**
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The measure used the PRISM III algorithm, a proprietary risk adjustment scheme.
- Information about the model has been published previously.
- The developer states elsewhere (measures #0334, #0335) that for the VPS system, "numerators, denominators and all definitions are standardized with an inter-rater reliability (IRR) >96%." From this one could infer that validity testing at the data element level was conducted.
  - Per NQF guidance, separate reliability testing is not required when validity testing at the data element level is performed for all critical data elements.
  - The following section addresses whether the Committee wishes to consider whether data element-level validity testing was performed.

Describe any updates to testing:

- The developer attests no new testing was performed for this submission.

SUMMARY OF TESTING
Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Results of reliability testing:
- The results of the data element level validity testing are provided in the following section.

Guidance from the Reliability Algorithm: Not applicable

Questions for the Committee:
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.
Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:
- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The measure uses the PRISM III algorithm, a proprietary risk adjustment scheme.
- Information about the model has been published previously.
- The developer attests no new testing was performed for this submission.
- The developer's previous submission stated "SMR is an established method which has been demonstrated to be reliable and valid. Further testing is not indicated."

Describe any updates to validity testing:

- The developer attests no new testing was performed for this submission.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer's previous submission stated "SMR is an established method which has been demonstrated to be reliable and valid. Further testing is not indicated."
- Elsewhere for its other related measures (#0334, #0335), the developer states "since numerators, denominators, and all definitions are standardized with an IRR >96%, this variation reflects differences in care and not the measurement itself." From this one could infer validity testing at the data element level was assessed.

Validity testing results:

- The developer states elsewhere for the VPS system that "numerators, denominators and all definitions are standardized with an inter-rater reliability (IRR) >96%." From this one could infer validity testing at the data element level was assessed.
  - The developer does not provide any additional information on the reliability of specific data elements.
  - Only reference to inter-rater reliability is made. The developer does not provide additional analyses (e.g., sensitivity, specificity, PPV, NPV).

Questions for the Committee:

- *Does the Committee wish to discuss with the developer whether the IRR data cited for its other measures apply here?*
- *The developer attests the specifications have not changed, and no new testing has been conducted. The developer provides the citation for the PRISM III model considered for the last endorsement review. Does the Committee agree is no need for repeat discussion and vote on Validity (and Reliability, if the Committee accepts the literature cited and/or the IRR data)?*

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following information:

- The endorsed severity adjustment methodology used for calculating SMR excludes:
  - PICU patients >=18 years of age
  - PICU patients under the age of 18 years with a stay < 2 hours in the PICU or < 2 consecutive sets of vital signs consistent with life
  - Patients admitted to PICU for palliative care
    - Palliative cases are excluded because the intention is that the patient will likely die in the ICU, skewing SMR calculations if the patients are included.
  - Preterm infants post-gestational age 36 weeks
- The other exclusions are consistent with the use of the PRISM 3 instrument for severity adjustment. The tool has not been validated in patients <36 weeks gestation, > or equal to 18 years, or if not in the PICU at least two hours/for two vital signs to be taken.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**
SDS factors included in risk model? ☐ **Yes** ☒ **No**

The developer provides the following information:

- As with the 2012 version of this measure the risk adjustment model used is proprietary.
- The risk model was developed using forward stepping logistic regression. Final variables were selected using a significance level p<0.05.
- The developer attests the specification have not changed since the previous review. The risk factor variables used in the version of PRISM 3 currently in use in the VPS dataset include:
  - PRISM 3 12-hour score
  - PRISM 3 12-hour score squared
  - Pre-ICU care area
  - Operative status
  - Acute diagnosis of diabetes
  - Pre-ICU cardiac massage
  - Age
- No statistically significant differences for the SMR was found for the following race/ethnic groups: 'African American', 'American Indian/Indigenous', 'Asian/Indian/Pacific Islander', 'Caucasian/European Non-Hispanic', 'Hispanic', and 'Other/Mixed'. Race and ethnicity is reliably collected by 85% of sites.
- No statistically significant differences for the SMR was found for the following insurance statuses: 'Managed Care', 'Commercial/Indemnity Insurance', Medicaid & Medicaid Managed Care', 'Self-

pay', and 'Other'. Primary payer is reliably collected by 38% of sites.

- The developer validated the risk model as described for the previous submission.
  - Training sample
    - Used 7,375 admissions including 213 deaths. The model predicted 213.002 deaths (SMR = 1.00, z = 0.000, p = 1.0).
    - The developer reported that area under the ROC curve, used to assess discrimination, was excellent with c = 0.930 +/- 0.009. Discrimination was further assessed using the misclassification rate for deaths, with a cutpoint for classifying deaths predicted to live assigned arbitrarily at p = 0.5. Using this method, the misclassification rate for deaths was 0.657.
    - The developer reported that the Hosmer-Lemeshow statistic, used with seven strata of mortality (0.00-0.01, >0.01-0.035, >0.035-0.075, >0.075-0.150, >0.150-0.250, >0.250-0.500, and >0.500), selected to ensure sufficient data in each stratum, demonstrated excellent calibration (H-L statistic = 5.4179, p = 0.36703).
  - Test sample
    - 3,208 admissions, including 93 deaths, were included in the test sample. Overall, 97.150 deaths were predicted (SMR = 0.957, z = -0.542, p = 0.583).
    - The developer noted that discrimination, as assessed by the area under the ROC curve, was excellent (c=0.944 +/- 0.013). The misclassification rate for deaths was 0.591 using the same method as used in the training sample.
    - The developer also reported that the Hosmer-Lemeshow test, using the same mortality strata as in the training sample, was excellent (HL = 7.205, p = 0.206).

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer provides the following information.

- VPS examined a cohort of 79 units that collected the entire year of 2014 data. That data set was comprised of 92,943 patients who were discharge from the participating pediatric ICUs within the calendar year of 2014. All of these patients meet the PRISM 3 inclusion criteria as defined by the author of the PRISM 3 tool. The number of cases by site range between 252 and 3284 with a median of 1072. The data include an indicator of outcome (ii.e., death before ICU discharge or transfer) and the PRISM 3 calculated risk of mortality/probability of death score.
- A calculation of SMR at individual unit level with comparison was conducted. SMRs are compared to a z-test for statistical significance. This large sample approximation test allows for customized cohorts for comparison and equivalence testing.
- The unit-level standardized mortality ratio (SMR) ranges between 0.16 and 2.02. Over 2014, the median unit-level SMR was 0.92 with interquartile rate of 0.79 - 1.15. The mean unit-level SMR was 0.97 with standard deviation of 0.32. The patient-level mean SMR for 2014 is not statistically different from 1.0 (0.99; 95% CI, 0.96 - 1.02)

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not Applicable

2b7. Missing Data

- Missing data analysis was not conducted on this measure

Guidance from the Validity Algorithm: 1→2→3→10→11 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The specifications are consistent with the evidence. The target population is clearly defined. No concerns with validity other than what I described in the reliability discussion above.

**Agree that the score from this measure is a reasonable measure of quality.

**Specs are c/w the evidence but, based upon pre-selection bias, the fact that a particular pt. died when PRISM III scoring did not predict that certainly might relate to poor care but might not reflect a lack of quality care by the PICU. Other measures such as LOS, un-preplanned readmission rate, line infections, skin breakdown, VAPs, etc., may be more useful direct measures of QOC.

2a2. Reliability Testing

Comments:

**There is some concern that the severity of patient mix may not be adequately accounted for in the methodology leading to potential inaccurate results when reporting outcomes. This is particularly relevant when the developer indicated that when measured over years, there was little change in outcomes.

**Comfortably with validity testing as noted in the measure worksheet

**It is presumed that testing was conducted at both data element and score levels. I continue to have doubts about the premise - that, if patients die in your PICU at an actual rate higher than a predicted rate (which is calculated using proprietary software and black-box scoring), it must stem from a quality of care deficit in some aspect(s) of care. While it "makes sense," in some way, by not pointing to any specific deficit, it seems arbitrary and unhelpful.

2b2. Validity Testing

Comments:

**Administrative data is used for this measure. Therefore there should not be a concern with missing data or no response results.

**not likely as variables are discrete data elements not subject to interpretation.

**Exclusions are valid, laterally, but leaving out premies from the mix is just unrealistic, particularly in some units. There is a conceptual relationship between measure and focus. Some variables may not be present at the start of care, such as intraoperative complications, post-operative decompensation, etc. Yes, risk adj approp developed, except as noted. Probably analyses indicate acceptable results. Yes, approp risk-adj strategy. Measure does not identify specific clinically-meaningful, practical differences with respect to quality, i.e., "actionable." Results support the risk-adj approach. Missing data could skew the results of a small-volume PICU.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:

**The developer did not conduct any new reliability tests on the measure. It relied on the inter-rater reliability testing from the original submission.

**Reliability reasonable on earlier testing.

**Reliability testing suggests that the data elements are repeatable over an extended time and either are precise enough to show that there is not even any trending going on or, in reality, so vague as to yield generalized results ("regression to the mean").Given comments regarding the IRR and low variability, further validity testing is suggested but I would like to hear more from the developer on this.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer provides the following information:

- Some data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. All necessary data may be available electronically if an organization has implemented an EHR. In the absence of an EHR, the developer notes that manual data collection would be required.
- There are no fees or licensing required to collect and calculate this measure. There are potential fees associated with using available software products such as the VPS. However, VPS provides these data free of charge to participating member units. The 2016 VPS participation fee schedule consists of a baseline fee with incremental charges based on unit volume. Small units pay $16,590 per annum and the largest units pay $33,150 per annum; the VPS fee schedule attached in addendum A1.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments:

**The data is available and would be routinely calculated or collected during care delivery. Providers may utilize different tools to determine risk for mortality which may yield varying results. Also transfer of patients to different floors/units may make the data more challenging to collect.
**Only concern is as a proprietary measure.
**Predicted mortality rate is not routinely generated, calculated, or used during care delivery. All necessary data might be available electronically but the "scoring" is proprietary and access to scoring essentially requires a membership fee. Operationally: EMR + $$ = access to scoring OR no EMR + manual review + data entry + $$ = access to scoring. Without identification of specific aspects of care which to target, however, I am not sure how practical this would be.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- The measure data are not aggregated and publicly reported; however, some hospitals (e.g., in Wisconsin, Texas) participating in the VPS system may individually publicly report their data.
- The developer states that the funding body for California pediatric healthcare, California Children's Services, mandates use for accountability purposes.
- VPS currently collects data from 79 hospitals, which use the measure to benchmark against a VPS reference group of other participants.

Improvement results:

- From 1/1/2012 until 12/31/2014 the quarterly SMR for VPS's U.S. PICUs (PICUs only) has varied between 0.93-1.02 (n=273,130). No increasing or decreasing trend for the overall SMR (Spearman correlation test, p=0.39).

Unexpected findings (positive or negative) during implementation:

- The developer states there were no unexpected findings during implementation.

Potential harms:

- The developer reports there were no identified unintended consequences for this measure during testing or since implementation.

Feedback: No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Is use only in local programs sufficient? Should the measure be implemented on a broader scale?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**The developer indicates that the data is publicly reported, yet the narrative indicates that it is only publicly reported by a few facilities that use the developers tools/system.

**Publicly reported seems vague. It seems to imply that some hospitals are using the data to post on their external web sites, is this public reporting? Is California reporting enough to satisfy the measure?

**The measure is currently only being voluntarily publically reported and, presumptively, only when results are positive and self-promoting. Data has not yet resulted in trending - "bending" the mortality curve. Without pointing a poor-performing unit in a specific direction(s), whereby they could affect changes in care processes, in order to reduce unexpected mortality, I am not sure of the utility of the measure as a QI tool.

Criterion 5: Related and Competing Measures
Related or competing measures • No related or competing measures identified

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0343 NQF Project: Pulmonary Project

1. IMPACT, OPPORTUITY, 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)

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 specifically relates to a health outcome- mortality.

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

Selected individual studies (rather than entire body of evidence)

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)*:

Standardized mortality ratio is a well established methodology for comparing two or more care settings while adjusting for patient level case-mix data. Consequently, there is no perceived need in the PICU community for SMR re-evaluation aside from periodic recalibration of the tool used for risk adjustment. Hospital SMR has been an important measure for evaluating outcomes.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: A search of Pubmed returned 155 articles specific to using SMR in an intensive care unit setting.

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): The standardized mortality ratio is simply a ratio. It is accepted as an appropriate quality measure for ICU settings. The literature is consistent on the value of using SMR with three caveats. First, use of a calibrated tool for severity adjustment has been identified as important. Second, a recent publication by Brinkman in Critical Care Medicine (2012; 40(2)373-378) identified that use of a physiology-based tool when calculating SMR is superior than using administrative data. Finally, it is important to note that premature transfer of patients from an ICU can lower the SMR, creating a potential for "gaming" measures (Kahn Chest. 2007;131(1):68-75). Use of measure 343 in conjunction with measure 334 and 335 address the potential for gaming.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* SMR has been found to be a reliable and valid tool for measuring ICU quality. Individual ICUs have demonstrated a decrease in SMR as a result of targeted improvement efforts.

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

SMR is beneficial in comparing outcomes of otherwise dissimilar PICUs while allowing individual organizations to trend their own performance.

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

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

1c.12 If other, identify and describe the grading scale with definitions: per 1c.9 above no grading has been done.

1c.13 Grade Assigned to the Body of Evidence: N/A

1c.14 Summary of Controversy/Contradictory Evidence:

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

Brinkman S, Abu-Hanna A van der Veen A. A comparison of the performance of a model based on

administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ration of intensive care units. Critical Care Medicine (2012; 40(2)373-378

Kahn JM, Kramer AA, Rubenfield GD. Transferring critically ill patients out of hospital improves the standardized mortality ratio: a simulation study.

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

N/A

1c.17 Clinical Practice Guideline Citation:

1c.18 National Guideline Clearinghouse or other URL:
http://qualitymeasures.ahrq.gov/content.aspx?id=33605&search=standardized+mortality+ratio

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: per 1c.19 above no grading has been done.

1c.23 Grade Assigned to the Recommendation:

1c.24 Rationale for Using this Guideline Over Others:

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: Moderate 1c.26 Quality: Moderate1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0343_Evidence_MSF5.0_Data_-2015.123015.-635876261896384392.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Use of SMR allows for comparing care in different PICUs while allowing for accounting for variations in the severity of patients´ illness. While there is no "right" mortality rate for a PICU, comparisons of SMR can guide quality improvement efforts.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Current performance:
The data from 79 PICUs submitting PRISM III data for patients discharged in 2014. In total, 92,943 individual PICU discharges and transfers were recorded.
The unit-level standardized mortality ratio (SMR) ranges between 0.16 and 2.02. Over 2014, the median unit-level SMR was 0.92 with interquartile rate of 0.79 – 1.15. The mean unit-level SMR was 0.97 with standard deviation of 0.32.

The patient-level mean SMR for 2014 is not statistically different from 1.0 (0.99; 95% CI, 0.96 – 1.02).

Performance overtime:
An analysis of 273,130 cases discharged between 1/1/2012 and 12/31/2014 showed no monotonic trend (i.e. no increasing or decreasing trend) for SMR by quarter (Spearman's correlation test p=0.39).

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Population differences have not been found to be variable in pediatric intensive
care therapies. A study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status. After adjustment for differences in illness severity, standardized mortality ratios and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care unit.

Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.

http://pediatrics.aappublications.org/content/127/3/e588.short
http://link.springer.com/article/10.1007/s10900-014-9823-0#/page-1

VPS stratifies by race/ethnicity, age groups, gender, and insurance payer.

Except for two categories, no statistically significant differences for SMR were observed for the age categories ('<1 Month', '1 Month - 23 Month', '2 Years - 5 Years', '6 Years - 12 Years', '13 Years - 18 Years'). First, the age group '1 month-23 months' (n=27,732) had a higher SMR (1.08; 95% CI, 1.02 - 1.14, p=0.0002) than all other age groups. Second, the age group '13 years-18 years' (n=21,445) had a lower SMR (0.89; 95% CI, 0.82 - 0.96, p=0.0021) than all other age groups.

No statistically significant differences for the SMR was found for the following race/ethnic groups: 'African American', 'American Indian/Indigenous', 'Asian/Indian/Pacific Islander', 'Caucasian/European Non-Hispanic', 'Hispanic', and 'Other/Mixed'. Race and ethnicity is reliably collected by 85% of sites.

No statistically significant differences for the SMR was found for the following insurance statuses: 'Managed Care', 'Commercial/Indemnity Insurance', Medicaid & Medicaid Managed Care', 'Self-pay', and 'Other'. Primary payer is reliably collected by 38% of sites.

No statistically significant differences for the SMR was found between gender/sex.

The higher SMR in lower age groups may reflect physiologic differences or vulnerability of this population related to their dependancy on care givers pre-hospitalization. Similarly the lower SMR in the teen-aged population may reflect physiologic differences.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Lopez A, Tilford J, Anand K, et al. Variation in pediatric intensive care therapies and outcomes by race, gender and insurance status. Ped Crit Care Med 2006; 7(1). 2-6.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

PICUs are a source of significant healthcare cost(1,2)
Risk adjusted mortality has been identified as a major focus of quality assessment in ICUs (6) and standardized mortality ratio (SMR)is recognized as a valid and reliable tool for comparing outcome between ICUs (3,4)
Quality of care in the PICU may be independent of indicators of overall hospital care. (5)

Therefore SMR remains a valid and reliable measure of ICU quality.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. Chalom R, Raphaely R, Costarino A. Hospital costs of pediatric intensive care. Crit Care Med; 1999. 27(10). 2079-2085.
2. Klem S, Pollack M, Geston P. Cost, resource utilization and severity of illness in intensive care. J Peds. 1990; 116(2). 231-237.
3. Knaus W, Wagner D, Zimmerman JE, Draper E. Variations in mortality and length of stay in intensive care units. Ann of Int Med. 1993;118(10).753-761
4.Selker H. Systems for comparing actual and predicted mortality rates: characteristics to promote cooperation in improving hospital care. Ann Int Med. 1993; 118228(10). 820-822
5. Pollack MM, Cuerdon TT, Patel, KM, et.al. Impact of quality-of-care factors on pediatric intensive care unit mortality. Journal
of the American Medical Association;1994:272 (12). 941-946.
6. Glance, L.G., Osler TM, Dick A. Rating the quality of intensive care units: Is it a function of the intensive care unit scoring system? Critical Care Medicine; 2002.30(9):1976-1982.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://s3.amazonaws.com/vpspublic/NQFMeasures.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
None, the risk model has not changed.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Actual number of deaths occurring in PICU.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
All PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment. Data submission quarterly with reporting on annual basis.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Exclusions:
- PICU patients >=18 years of age
- PICU patients under the age of 18 years with a stay < 2 hours in the PICU
- < 2 consecutive sets of vital signs consistent with life
- Patients housed in the ICU on boarder status or Intermediate care status

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The sum of of predicted PRISM 3 mortality. "Predicted mortality" = Number of deaths expected based on assessed physiologic risk of mortality.

Include all PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment or boarder/IMCU status.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Inclusions:
- All PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

Include all PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment or boarder/IMCU status.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
All PICU patients >= 18 years of age, PICU patients with a stay < 2 hours or < 2 consecutive sets of vital signs consistent with life, deaths occurring outside the PICU, patients admitted to PICU for palliative care: AAP Committee on Bioethics

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
No additional stratification occurs beyond the risk adjustment inherent to this measure. That is, the expected mortality that serves as the denominator in this measure specifically accounts for the severity of illness of patients included in the measure. No further stratification is appropriate based on current literature.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Selection criteria for risk adjustment tool for pediatric ICU's:
- Tool must allow quality assessment and comparison between intensive care units, and must be widely used
- Tool must be valid and reliable for severity adjustment and measurement of quality of care provided
- Computation of mortality risk must be in the public domain (i.e. free of charge)
- Algorithms must receive ongoing validation and recalibration

The PRISM 3 model meets these criteria.

The risk model was developed using forward stepping logistic regression. Final variables were selected using a significance level $p<0.05$.

The risk factor variables used in the version of PRISM 3 currently in use in the VPS dataset include:
- PRISM 3 12-hour score
- PRISM 3 12-hour score squared
- Pre-ICU care area
- Operative status
- Acute diagnosis of diabetes
- Pre-ICU cardiac massage
- Age

1. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996;24:743-52.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also*

indicate if available at measure-specific URL identified in S.1.)
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
As with the 2012 version of this measure the risk adjustment model used is proprietary.

S.16. **Type of score:**
Ratio
If other:

S.17. **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

S.18. **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.)*
PRISM 3 is a valid, realiable and internationally accepted risk measurement tool. The methodology and measure specifications have been published(1) and are available at
https://s3.amazonaws.com/vpspublic/NQFMeasures.pdf

1. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996;24:743-52.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. All patients are included.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
None

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.

No mandatory data source or collection instrument for PICU community. Potential resources include PICU-specific databases or the VPS database (myvps.org).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0343_MeasureTesting_MSF5.0_Data_2016.0114..doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0343 NQF Project: Pulmonary Project

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.

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)*:

SMR is an established method which has been demonstrated to be reliable and valid. Further testing is not indicated.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*: N/A

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

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:

Exclusion criteria for the severity adjustment tool (PRISM 3) assures accuracy of SMR calculation.

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)*:

SMR is an established method which has been demonstrated to be reliable and valid. Further testing is not indicated.

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

N/A

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)*: N/A

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)*:

The endorsed severity adjustment methodology used for calculating SMR excludes:

• PICU patients >=18 years of age

• PICU patients under the age of 18 years with a stay < 2 hours in the PICU or < 2 consecutive sets of vital

signs consistent with life

• Patients admitted to PICU for palliative care

• Preterm infants post-gestational age 36 weeks

Palliative cases are excluded because the intention is that the patient will likely die in the ICU, skewing SMR calculations if the patients are included.

The other exclusions are consistent with the use of the PRISM 3 instrument for severity adjustment. The tool has not been validated in patients <36 weeks gestation, > or equal to 18 years, or if not in the PICU at least two hours/for two vital signs to be taken.

Further validation of these exclusions is not indicated.

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

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

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)*:

PRISM 3 is a valid, reliable and internationally accepted risk measurement tool. The methodology and measure specifications have been published(1) and are available at https://s3.amazonaws.com/vpspublic/NQFMeasures.pdf.

Calibration reassessment has been performed with plans for future model enhancement PRISM 3 is a valid, reliable and internationally accepted risk measurement tool. The methodology and measure specifications have been published(1) and are available at https://s3.amazonaws.com/vpspublic/NQFMeasures.pdf.

See 2b4.2 for additional information on analytic method.

Calibration reassessment has been performed with plans for future model enhancement

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

Methods

Training sample

The training sample used 7375 admissions including 213 deaths. The model predicted 213.002 deaths (SMR = 1.00, z = 0.000, p = 1.0).

Discrimination

The area under the ROC curve, used to assess discrimination, was excellent with c = 0.930 +/- 0.009. Discrimination was further assessed using the misclassification rate for deaths, with a cutpoint for classifying deaths predicted to live assigned arbitrarily at p = 0.5. Using this method, the misclassification rate for deaths was 0.657.

Calibration

The Hosmer-Lemeshow statistic, used with seven strata of mortality (0.00-0.01, >0.01-0.035, >0.035-0.075, >0.075-0.150, >0.150-0.250, >0.250-0.500, and >0.500), selected to ensure sufficient data in each stratum, demonstrated excellent calibration (H-L statistic = 5.4179, p = 0.36703).

Test sample

3208 admissions, including 93 deaths, were included in the test sample. Overall, 97.150 deaths were predicted (SMR = 0.957, z = -0.542, p = 0.583).

Discrimination

Discrimination as assessed by the area under the ROC curve was excellent (c=0.944 +/- 0.013). The misclassification rate for deaths was 0.591 using the same method as used in the training sample.

Calibration

The Hosmer-Lemeshow test, using the same mortality strata as in the training sample was excellent (HL = 7.205, p = 0.206).

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):

See 2b4.2 for additional information on analytic method including calibration and discrimination results.

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)*: A population of 117 PICUs using the VPS database found SMR varies widely. According to the latest data through September 2011 available in the VPS, the average SMR ranged from 0.00 to 1.76.

VPS examined a cohort of 79 units who collected the entire year of 2014 data. That data set was comprised of 92,943 patients who were discharge from the participating pediatric ICUs within the calendar year of 2014. All of these patients meet the PRISM 3 inclusion criteria as defined by the author of the PRISM 3 tool. The number of cases by site range between 252 and 3284 with a median of 1072. The data include an indicator of outcome (ie: death before ICU discharge or transfer) and the PRISM 3 calculated risk of mortality/probability of death score.

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

This was merely a calculation of SMR at individual unit level with comparison. SMRs are compared with a z-test for statistical significance. This large sample approximation test allows for customized cohorts for comparison and equivalence testing.

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

The unit-level standardized mortality ratio (SMR) ranges between 0.16 and 2.02. Over 2014, the median unit-level SMR was 0.92 with interquartile rate of 0.79 - 1.15. The mean unit-level SMR was 0.97 with standard deviation of 0.32.
The patient-level mean SMR for 2014 is not statistically different from 1.0 (0.99; 95% CI, 0.96 - 1.02)

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)*:

Published literature indicates use of administrative data is inappropriate for determination of SMR when clinical data is available. Thus, this is a clinical measure.

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

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)*: N/A

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)*: N/A

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

2.1-2.3 Supplemental Testing Methodology Information: None

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. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**
All necessary data may be available electronically if an organization has implemented an EHR. In the absence of an EHR, manual data collection would be required.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
The work required in collecting those elements necessary for calculating expected mortality (SMR denominator) is not insignificant but quite feasible. For instance, the group of 79 PICUs using the VPS database have collected these elements for more than 92,943 patient encounters between Q 1 and Q 4 2014.

Finally, the elements needed for determining the SMR denominator are also used in NQF measure 0334

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees or licensing to collect and calculate this proportion. There are potential fees associated with using available software products such as the VPS.There are no fees or licensing to collect and calculate this proportion. There are potential fees associated with using available software products such as the VPS. (VPS fee schedule attached in addendum A1.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Children´s Hospital Wisconsin http://www.chw.org/ Cleveland Clinic Children´s Hospital http://my.clevelandclinic.org/ccf/media/files/outcomes/2013/outcomes-peds.pdf Texas Children´s Hospital http://www.texaschildrens.org/About-Us/Quality-Measures/Pediatric-Intensive-Care-Unit/ Payment Program California Children´s Health Services http://www.dhcs.ca.gov/services/ccs/Pages/default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) California Children´s Health Services http://www.dhcs.ca.gov/services/ccs/Pages/default.aspx Quality Improvement (Internal to the specific organization) 79 hospitals have received this data Re to: QI initiatives https://myvps.org

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

VPS currently provides data to 79 U.S. PICUs which use the measure to benchmark against a VPS reference group of other participants. This measure is used for internal Quality Improvement at the individual site level. Many of these facilities report patient safety data. An example of three of the participating hospitals listed above list this benchmark data on their hospital-specific web sites.

California Children´s Health Services(CCHS/CCS) includes 28 California based PICUs which participate in VPS and contribute data specific to this measure. In 2014, 14 of the 28 sites contributed data by the California deadline resulting in a state specific performance report to California Children´s Services.

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment

program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

From 1/1/2012 until 12/31/2014 the quarterly SMR for VPS's U.S. PICUs (PICUs only) has varied between 0.93 - 1.02 (n=273,130). No increasing or decreasing trend for the overall SMR (Spearman correlation test, p=0.39).

Geographic area update: 44 states, 145 PICU/CICUs

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
The use of SMR is a well described and validated means of compared patient relevant outcomes. The fact that the aggregate SMR approaches 1.0 suggests that pediatric ICUs in this data cohort do not have higher than expected mortality. Therefore, improvement would not necessarily be expected. If however, the trend in SMR overtime reveals an increase then improvement would be in order. Consequently SMR should remain an endorsed measure as a means for ongoing monitoring of outcomes.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Manual data abstraction with entry into a multi-institutional clinical PICU database (the VPS (myvps.org) has been completed for the variables used in this measure since 2002. Currently, 122 PICUs are abstracting and entering data with an aggregate interrater reliability of 96.81%

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** VPS_Fee_Schedule_CY16-635884097727308239.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Virtual PICU Systems, LLC

Co.2 **Point of Contact:** Nancy, Brundage, nbrundage@myvps.org, 888-999-4850-103
Co.3 **Measure Developer if different from Measure Steward:** Virtual PICU Systems, LLC
Co.4 **Point of Contact:** Mat, Scanlon, mscalon@mcw.edu, 888-999-4850-

Additional Information

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

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 12, 2015
Ad.4 **What is your frequency for review/update of this measure?** 3 years
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** The severity-adjusted mortality rate uses a proprietary adjustment tool.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0468
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: The measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR). Mortality is defined as death for any cause within 30 days after the date of admission for the index admission, discharged from the hospital with a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary diagnosis of pneumonia (including aspiration pneumonia) coded as present on admission (POA). CMS annually reports the measure for patients who are 65 years or older and are either Medicare fee-for-service (FFS) beneficiaries and hospitalized in non-federal hospitals or patients hospitalized in Veterans Health Administration (VA) facilities.

Please note this measure has been substantially updated since the last submission; as described in S.3., the cohort has been expanded. Throughout this application we refer to this measure as version 9.2.
Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality 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 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.

Pneumonia mortality is a priority area for outcomes measure development as it is an outcome that is in part attributable to care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices, as well as increase transparency for consumers.

Numerator Statement: The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis.
Denominator Statement: This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or over or (2) patients aged 18 years or older. We have specifically tested the measure in both age groups.

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA but no secondary discharge diagnosis of severe sepsis; and with a complete claims history for the 12 months prior to admission. The measure will be publicly reported by CMS for those patients 65 years or older who are Medicare FFS beneficiaries admitted to non-federal hospitals or patients admitted to VA hospitals.

Additional details are provided in S.9 Denominator Details.

Denominator Exclusions: The mortality measures exclude index admissions for patients:

1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility;
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
4. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Mar 09, 2007 **Most Recent Endorsement Date:** Jul 31, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

The developer reports the following:

- This measure calculates hospitals' 30-day risk-standardized mortality rate for patients who have been hospitalized with pneumonia.
- As a rationale for measuring this health outcome, the developer states hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education.
- The developer reports studies demonstrate appropriate, guideline-recommended care and timely treatment for pneumonia patients can reduce the risk of mortality within 30 days of hospital admission.

Guidance from the Evidence Algorithm : 1→2 (eligible for PASS rating)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for

improvement.

- The previous Committee noted that no significant change had been seen in the 3-4 years of data provided (the mean for 2007 to 2009 period was 11.7%, with a range of 6.9% to 20.4%), but because CMS combines 3 years of data for public reporting, the Committee concluded it was too soon to expect significant change.
- The developer reports aggregate performance data, as follows:

	7/2011 – 6/2012	**7/2012 – 6/2013**	**7/2013 – 6/2014**	**7/2014 – 6/2015**
# hospitals	4,614	4,605	4,566	4,694
#admissions	460,836	482,891	434,262	1,377,989
Mean rate (SD)	16.8 (1.6)	16.7 (1.6)	15.5 (1.4)	16.4 (2.0)
Range (min-max)	11.0-23.9	11.1-25.7	10.5-22.8	8.7-25.4
10th percentile	15.0	14.8	13.9	14.0
90th percentile	18.8	18.7	17.3	18.9

Disparities

- The developer provides the following information (July 2011-June 2014):

	#hospitals	***# admissions***	***Minimum rate***	***Median rate***	***Maximum rate***
Dual eligibles					
High proportion	1,083	249,621	8.7	16.3	25.4
Low proportion	1,081	376,625	10.7	16.0	24.6
African Americans					
High proportion	1,082	399,018	24.5	16.2	10.9
Low proportion	1,252	145,085	23.4	16.3	8.7
AHRQ SES score					
High proportion	1,082	218,219	25.4	16.6	10.4
Low proportion	1,082	347,875	22.7	16.2	10.7

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *The 3-year median performance for the last evaluation was 11.7%, and the developer reports a 3-year performance for 2011-2014 of 16.4%. Does the Committee wish to discuss this change in gap with the developer?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**• This measure calculates hospitals' 30-day risk-standardized mortality rate for patients who have been hospitalized with pneumonia.

• As a rationale for measuring this health outcome, the developer states hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education.

• The developer reports studies demonstrate appropriate, guideline-recommended care and timely treatment for pneumonia patients can reduce the risk of mortality within 30 days of hospital admission.

**The outcomes measure evidence relates to the outcomes being measured. There is a direct correlation of interventions to outcomes.

**The outcomes is supported by at least one intervention.

**As a rationale for measuring this health outcome, the developer states hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education. I agree that the developer has sufficient evidence to support the measure, however the implementation challenge lies in providers having the processes in place and ability to track

those activities to determine where improvement is needed at their respective organizations and the extent of impact when changes are made. How does one determine what is the degree of influence/impact that the aforementioned activities have on mortality rates (e.g., discharge planning reduced by 5% whereas med rec 10% etc., so aiming for changes in activities with the potential for greatest impact on outcomes)

Not clear on rationale behind adding aspiration pneumonia to the broadened definition of pneumonia. Clear information was provided as to the need to use a PDx of sepsis and a secondary of pneumonia. It would be helpful to better understand inclusion of that as well.

**Sure. Early, appropriate antibiotics.

Without a change, there is probably no reason to revisit. However, do wonder about the use of this measure in patients 18-65 as mortality, use of hospice, and pre-existent conditions are likely to be much harder to assess in this group.

1b. Performance Gap

Comments:

**Yes. It shows discrepancy gaps between the upper and lower proportions in all examples and most apparently in the African American proportions.

**A gap was demonstrated in the data for the dual eligible population where interventions may improve outcomes.

**Yes some was reported. It is not clear that it demonstrates a gap in care. However, when the disparity date is included, it does appear that some gaps may exist and some disparities may exist as well.

**The 3-year median performance for the last evaluation was 11.7%, and the developer reports a 3-year performance for 2011-2014 of 16.4%. It seems that mortality rates are increasing:

1) Were there any changes in measure definitions and specifications that resulted in an

2) If not, has the developer looked into what is happening in the healthcare landscape that is negatively impacting?

There are definitive opportunities for improvement with rates largely unchanged through reported data 2011-2015, however with that many years and not much improvement, nationally or any clear disparities, there's a potential need to outline potential underlying processes that can improve hospital performance scores

Additionally, how will the broadening of the PNA diagnosis impact rates?

The following is not a developer issue, but rather a CMS consideration-- In terms of reporting, CMS' rolling 3-years of data will include differing data definitions, considering how the expanded pneumonia diagnosis impacts rolling the score up with differing definitions

**The median rate seems to be the same across all "vulnerable" populations.

Yes. There has now been an increase in mortality. Seems unlikely that the measure is leading to real improvements in care.

1c. High Priority (previously referred to as High Impact)

Comments:

**n/a

**n/a

**NA

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

- In addition to routine updating of the condition categories used in the risk-adjustment approach, the patient cohort included in the measure has been expanded to include patients with a principal discharge diagnosis of aspiration pneumonia and those with a principal discharge diagnosis of sepsis (not including severe sepsis) who have a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as present on arrival and no secondary discharge diagnosis of severe sepsis. The expansion to include patients with a sepsis discharge diagnosis was recommended via public comment during the last NQF evaluation of this measure in

2012.

- The measure is specified as a facility-level measure for the hospital/acute care setting; the data source for the measure is administrative claims.
- This measure can be used in either of two patient cohorts: (1) patients 65 years or older, or (2) patients 18 years or older.
- Specifications include ICD-9 and ICD-10 codes to identify patients with a primary discharge diagnosis of pneumonia or primary diagnosis discharge of sepsis with secondary diagnosis of pneumonia, date of birth, and transfer status, based on admission and discharge dates (used in denominator) and discharge disposition (used for exclusions). No specific code for identifying death is provided, as the developer has used different datasets (Medicare Enrollment Database; California vital statistics file) in testing of this measure. The developer also notes that other data sources (e.g., the CDC's National Death Index or the SSA's Death Master File could be used to identify date of death).
- ICD-10 codes are included in the specifications. The ICD-9, and ICD-10 codes are described in the numerator and denominator details, and included in the data dictionary attachment. The ICD 9→10 conversion goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
- Only those patients enrolled in both Part A and Part B Medicare for 12 months prior to the index admission are included when the measure is used for the Medicare population. It is unclear, although assumed, that there is a similar 12-month enrollment requirement for an all-payer population.
- If a patient has more than one index admission, only one is randomly chosen for use in the measure.
- Patients are excluded from the measure if age is > 115, the discharge date is prior to the admission date, sex is neither male nor female, admitted to hospice in the 12 months prior to the index admission or on the first day of the index admission, if discharged against medical advice, or if discharged alive on day of or day following index admission but not transferred to another acute care facility.
- The calculation algorithm, included in S.18, describes how the risk-standardized mortality ratio is calculated.
- This outcome measure is risk-adjusted using a statistical risk model with 36 factors (age, gender, and various co-morbidity indicators).

Questions for the Committee:

- *Are the changes in the specification appropriate? That is, can the expanded pneumonia cohort be reliably coded and captured?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The developer conducted reliability testing at the measure score level.

Describe any updates to testing: Reliability testing at the measure score level was conducted for a more recent time period (2011-2014)

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- Data used for reliability testing included July 2011-June 2014 Medicare fee-for-service inpatient and outpatient

claims; this dataset included information for 1,377,989 admissions and 4,697 hospitals hospitals.

- Developers used a split-sample (or "test-retest") methodology to test score-level reliability. This is an appropriate method. For this analysis, developers randomly assigned half of the patients in each hospital to two separate groups, calculated the performance measure score for each hospital in each of the two groups, and compared the agreement between each hospital's paired scores using the intra-class-correlation coefficient (ICC) and applying a correction factor to account for the overall sample size. The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospitals.

Results of reliability testing:

- The ICC values from the split-sample analysis 0.79, indicating that 79% of the variance in scores are due to differences between hospitals. According to the Landis and Koch classification, an ICC value of 79% can be interpreted as substantial agreement and a value of 0.7 is often regarded as a minimum acceptable reliability value. Note that the developer states that reliability testing was "was limited to hospitals with 12 or more cases in each split sample", although the measure itself is not limited to hospitals with 12 or more cases..

Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH rating)

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

- As noted earlier, the specifications were revised since the last NQF endorsement review, so that the pneumonia cohort includes those with a principal discharge diagnosis of aspiration pneumonia and those with a principal discharge diagnosis of sepsis (not including severe sepsis) who have a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis. The developer provides a rationale for this change and notes it is supported by the literature.

Question for the Committee:

- *Are the specifications, including the expansion of the patient population, consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The measure was originally validated by correlating the claims-based performance score results to results from a similar mortality measure that used clinical data obtained via manual chart audit of medical records for the same patient population. The findings of this analysis indicated that the two measures were highly correlated. An earlier version of the risk-adjustment model has been validated in a 2006 California all payer (18 years and older) dataset as well as the Medicare dataset.

Describe any updates to validity testing

- The developers also conducted additional testing of the risk-adjustment model using an updated dataset. However, no additional validity testing of the re-specified measure itself was conducted.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

The developer provides the following:

- In addition to the correlation analysis of an earlier version of the measure (described above), the only other empirical testing described by the developer included analyzing two sets of risk-model discrimination statistics after splitting a sample of a 2011-2014 Medicare fee-for-service dataset into two datasets.
- While this method does not establish completely the validity of the revised measure at the measure score level, it can help to validate the risk-adjustment approach.

Validity testing results:

- The model discrimination statistics (c-statistics) from the two from the split samples were 0.724 and 0.727. The developer notes that these are similar to the c-statistics found for the risk models of other mortality measures (e.g.., 0.72 for the initial model development and validation samples and 0.759 in the 2006 all-payer dataset; these data were reported in the previous submission of the measure).

Questions for the Committee:

- *Does the analysis of model discrimination provide enough additional validation of the measure, in the absence of empirical testing of the measure for the measure as currently specified? In other words, would you require any additional testing of the measure now that the patient population has been expanded?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- To ascertain impact of exclusions on the cohort, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion:
  - Discharged alive on the day of admission or the following day who were not transferred to another acute care facility - 3.73%
  - Age is > 115, discharge date prior to admission date, sex neither male nor female; date of death occurs before the date of discharge but the patient was discharged alive (*note: this last exclusion was not included in the specifications in section S.11*). – 0.00%;
  - Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission – 1.85%
  - Discharged against medical advice (AMA) – 0.54%

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

Description of the model

- This measure is risk-adjusted using hierarchical logistic regression model with 36 factors to create a hospital-level 30-day risk-standardized mortality rate that simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals.
- The model adjusts for case differences based on age, gender, and clinical status of the patient at the time of

admission (the latter using condition categories (CCs)).

- Only comorbidities that conveyed information about the patient at the time of admission or in the 12-months prior, and not complications that arose during the course of the hospitalization, were included in the risk-adjustment.
- After expansion of the patient cohort included in this measure, the risk-model was re-specified and now includes 5 additional condition code factors (septicemia/sepsis; disorders of fluid/electrolyte/acid-base; delirium and encephalopathy; respiratory dependence/tracheostomy; decubitus ulcer of skin); also, two condition categories (pneumonia; cardio-respiratory failure or shock) were expanded to include additional diagnoses.
    - Several datasets have been used in the development of this measure. The most recent update to the risk-adjustment approach uses Medicare claims from July 1, 2011 – June 30, 2014. The measure was also applied to a 2009 California all-payer database .
    - The table in section 2b4.4a lists the variables (and associated odds ratios) included the most recent version of the risk-adjustment model.

Performance of the model

- The *c*-statistic value computed using data from the 2011-2014 Medicare dataset was 0.716. The c-statistic is model discrimination statistic that represents the proportion of all-possible pairs with different observed outcomes for which the model correctly predicts a higher probability for observations with the outcome of interest than those without the outcome of interest. A *c*-statistic of 0.716 means that for 71.6% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable. .
- Developers also noted the predictive ability of the risk-adjustment model as another indicator of its discrimination: the lowest decile was 4.5% and the highest decile was 40.3%. A wide range between the lowest decile and highest decile indicates an ability to distinguish between high- and low-risk patients.
- The risk-decile plot based on the most recent dataset indicates good model fit (or calibration), as observed values are relatively similar to predicted values across the risk-deciles. Additional calibration statistics from the split Medicare samples also were provided [(γ0=-0. 0457, γ1=0. 9526) and (γ0=0496, γ1=9504), respectively]; these also indicate good model fit because the values of γ0 and γ1 are close to 0 and 1, respectively.
- A similar risk-adjustment approach was tested using a 2009 California all-payer dataset of patients ages 18+. The developers report good model discrimination and calibration from this approach and therefore consider use of the measure for all-payer patients ages 18+ to be appropriate.

Conceptual basis and empirical support for potential inclusion of SDS factors in risk-adjustment approach

- The developer noted that studies have shown either no or ambiguous racial disparities in pneumonia mortality and no literature on income or other SDS factors. However, they also stated that following SDS factors that have been examined in the mortality literature for various conditions: patient-level self-reported or documented race or ethnicity, income, and education level; median household; Agency for Healthcare Research and Quality (AHRQ)-validated SES index score; and the proportion of Medicaid patients served in the hospital. The developer identified several potential conceptual pathways to consider:
    - Relationship of socioeconomic status (SES) factors or race to health at admission.
    - Use of low-quality hospitals.
    - Differential care within a hospital.
    - Influence of socioeconomic status (SES) on mortality risk outside of hospital quality and health status.
- Based on their interpretation of the literature and analysis of the above pathways, the developers identified 3 potential SDS variables for potential inclusion in the risk-adjustment model:
    - African American race (as compared to all others)
    - Dual eligible status
    - AHRQ SES index score (based on 5-digit ZIP code data; includes percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room).
- Analyses indicate that the prevalence of these 3 SDS factors vary across measured entities and are associated with the measured outcome. However, when including any of the 3 SDS variables in a multivariable model that

includes all of the claims-based clinical variables, the effect size of each of these variables is small, the c-statistic is similar, and hospital-specific results change little. Moreover, the effect of each of the 3 SDS variable was protective, which is the opposite of what was expected based on the literature reviewed.
- Based on the empirical results, the developer decided NOT to include any of the 3 SDS factors analyzed in the final risk-adjustment model.

Questions for the Committee:
- *Is an appropriate risk-adjustment strategy included in the measure? That is, is the risk model adequate to control for difference in case mix across providers?*
- *Are all of the risk-adjustment variables present at the start of care?*
- *Do you agree with the developer's decision, based on its analysis, to not include SDS factors in the risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To demonstrate that meaningful differences between providers can be identified with this measure, the developer reports results as they are designated by CMS on the Hospital Compare website (i.e., based on a 95% interval estimate which could be lower or higher, or could include the national observed rate).
- These results indicate that out of 4,694 hospitals in the U.S., 244 performed "better than the U.S. national rate," 3,814 performed "no different from the U.S. national rate," and 271 performed "worse than the U.S. national rate." Another 365 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

Question for the Committee:
- *Does this measure identify meaningful differences about quality?*
- *Are the measure results meaningful to stakeholder audiences?*

2b6. Comparability of data sources/methods:

- Not Applicable

2b7. Missing Data

- An analysis of missing data analysis was not provided on this measure. Typically, however, there is very little missing data in claims measures.

Guidance from the Validity Algorithm : 1→2→3→6→7→8 (eligible for HIGH rating)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**I see no inconsistancies. The specifications were revised since the last NQF endorsement review, so that the pneumonia cohort includes those with a principal discharge diagnosis of aspiration pneumonia and those with a principal discharge diagnosis of sepsis (not including severe sepsis) who have a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis. The developer provides a rationale for this change and notes it is supported by the literature.

**The specifications demonstrate validity of the measure. The specifications are consistent with the evidence.

**n/a

**Not clear on what specific evidence supports inclusion of aspiration PNA in the expanded definition

**Yes

2a2. Reliability Testing

Comments:

**The measure was originally validated by correlating the claims-based performance score results to results from a similar mortality

measure that used clinical data obtained via manual chart audit of medical records for the same patient population. The findings of this analysis indicated that the two measures were highly correlated. An earlier version of the risk-adjustment model has been validated in a 2006 California all payer (18 years and older) dataset as well as the Medicare dataset.
The developers also conducted additional testing of the risk-adjustment model using an updated dataset. However, no additional validity testing of the re-specified measure itself was conducted.
**The validity testing was strong with a substantial number of entities and patients. Conclusions can be appropriately drawn based on the data. the score from this measure can be used as an indicator of quality.
**Validity was established by correlating the measure to a different measure of mortality that was based on chart audits. Prior versions had also been validated against all payer and Medicare data sets.
**Intent is unchanged in looking for PNA (just added the secondary PNA with primary sepsis); however what does addition of that do to qualifying cases? Seems that it'll "water down" the results thereby lowering the mortality rates. Would caution whether results from subsequent national data post-new specifications are due to this broadened definition of PNA vs. actual improvements on behalf of clinicians.
**As I understand the model discrimination test, they only looked at consistency (reliability) however they have not looked at validity as they did with the original model (The measure was originally validated by correlating the claims-based performance score results to results from a similar mortality measure that used clinical data obtained via manual chart audit of medical records for the same patient population.) As such, I think it is reasonable to ask for a similar assessment with these new defintions.
Overall, I would rate validity as moderate.
Yes, score can be specified as an indicator of quality (at least in part).

2b2. Validity Testing
Comments:
**The exclusions are consistant and there are no inappropriate exclusions.
room).
Analyses indicate that the prevalence of 3 SDS factors vary across measured entities and are associated with the measured outcome. However, when including any of the 3 SDS variables in a multivariable model that includes all of the claims-based clinical variables, the effect size of each of these variables is small, the c-statistic is similar, and hospital-specific results change little. Moreover, the effect of each of the 3 SDS variable was protective, which is the opposite of what was expected based on the literature reviewed. Based on the empirical results, the developer decided NOT to include any of the 3 SDS factors analyzed in the final risk-adjustment model.
- To demonstrate that meaningful differences between providers can be identified with this measure, the developer reports results as they are designated by CMS on the Hospital Compare website (i.e., based on a 95% interval estimate which could be lower or higher, or could include the national observed rate).

These results indicate that out of 4,694 hospitals in the U.S., 244 performed "better than the U.S. national rate," 3,814 performed "no different from the U.S. national rate," and 271 performed "worse than the U.S. national rate." Another 365 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
**No concerns.
**no threats seem to apply
**Would missing death date information be an issue here?
**2b3. Yes, exclusions are consistent with evidence. Not sure how to measure hospice in non-Medicare population.
I am not sure I understand why patients who are discharged on the day of admission or the following day are excluded. They state that they are unlikely to have "clinically significant" pneumonia. Do they have evidence to that effect? Evidence that the measure is a better assessment of quality when these patients are excluded?
2b4. Yes, risk adjustment seems appropriate.
Yes, all are present at the beginning.
If dual eligible have a 1% higher mortality, think we would want to include that as a factor. Using their tests for inclusion for these factors, how many of the other factors in the model would be added if they were not currently in the model?
2b5. Bell shaped curve. Not sure how that convincingly shows meaningful differences in quality.
Not sure how/if patients use the measure to assess hospital quality.

2b6. NA
2b7. NA
2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias

Comments:
**The developer conducted reliability testing at the measure score level.
Data used for reliability testing included July 2011-June 2014 Medicare fee-for-service inpatient and outpatient claims; this dataset included information for 1,377,989 admissions and 4,697 hospitals hospitals.
Developers used a split-sample (or "test-retest") methodology to test score-level reliability. This is an appropriate method. For this analysis, developers randomly assigned half of the patients in each hospital to two separate groups, calculated the performance measure score for each hospital in each of the two groups, and compared the agreement between each hospital's paired scores using the intra-class-correlation coefficient (ICC) and applying a correction factor to account for the overall sample size. The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospitals.
The ICC values from the split-sample analysis 0.79, indicating that 79% of the variance in scores are due to differences between hospitals. According to the Landis and Koch classification, an ICC value of 79% can be interpreted as substantial agreement and a value of 0.7 is often regarded as a minimum acceptable reliability value. Note that the developer states that reliability testing was "was limited to hospitals with 12 or more cases in each split sample", although the measure itself is not limited to hospitals with 12 or more cases.
**Reliability testing was performed in two methods. The results demonstrate an opportunity to generalize for widespread implementation. Testing was completed at the data element level.
**Reliability testing was done at the score level. In addition a split sample or test-retest methodology was used.
**The random assignment of an index admission when there are multiple relevant admissions poses challenges in consistency of implementation, however overall reliability testing at score level were sufficient without considering this.
**Relaibility: 4 Moderate. Yes, sufficient reliability.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications, including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.
**The data elements are administrative and electronic from the health record. There are not concerns with the data collection strategy.
**Feasibility should be high
**Concerns regarding death date and availability only in sources external to claims data.
**Low. Not sure why 30 day mortality would be considered to be "defined fields in electronic claims and generated or collected by healthcare personnel during the provision of care".

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- This measure is publically reported nationally in the Hospital Inpatient Quality Reporting (IQR) Program.
- The measure is used in the Hospital Value-Based Purchasing (HVBP) Program.

Improvement results : The developer reports aggregate performance data, as follows:

	7/2011 – 6/2012	**7/2012 – 6/2013**	**7/2013 – 6/2014**	**7/2014 – 6/2015**
# hospitals	4,614	4,605	4,566	4,694
#admissions	460,836	482,891	434,262	1,377,989
Mean rate (SD)	16.8 (1.6)	16.7 (1.6)	15.5 (1.4)	16.4 (2.0)
Range (min-max)	11.0-23.9	11.1-25.7	10.5-22.8	8.7-25.4
10th percentile	15.0	14.8	13.9	14.0
90th percentile	18.8	18.7	17.3	18.9

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms: There were no identified unintended consequences for this measure during testing or since implementation.

Feedback : No feedback provided on QPS. During the 2014-2015 MAP review, MAP conditionally supported this measure pending NQF review of the updates to this measure and continued endorsement.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**This measure is publically reported nationally in the Hospital Inpatient Quality Reporting (IQR) Program.
The measure is used in the Hospital Value-Based Purchasing (HVBP) Program.
Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.
Potential harms: There were no identified unintended consequences for this measure during testing or since implementation.
Feedback : No feedback provided on QPS. During the 2014-2015 MAP review, MAP conditionally supported this measure pending NQF review of the updates to this measure and continued endorsement.
The performance results can be used to further the goal of high-quality, efficient healthcare and the benefits of the measure outweigh any potential unintended consequences.
**The results of the use of the measure can be used to improve patient outcomes through outpatient interventions.
**The measure is intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes.
**Measure is useful, however largely impossible for hospitals to attempt to replicate and implement at institutional level to monitor performance and make more real-time advances to improve. Underlying processes shown to impact pneumonia mortality can be implemented and tracked but extent of impact on improving pneumonia outcomes would be difficult to discern outside of the infrequent CMS calculation and reporting of claims roll-up.

**Yes but am concerned by the few current groups who are using it.
Benefits do seem to outweigh unintended consequences.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0231 : Pneumonia Mortality Rate (IQI #20)

Harmonization

- The previous Committee determined #0231 and #0468 are related but not competing measures. Committee members felt strongly that both inpatient and 30-day mortality measures provide complimentary information and both are needed to describe the entire episode of care. However, the Committee asked whether further harmonization was possible and perhaps better alignment as they are both based on administrative data.
- The developer stated the pneumonia mortality measure cohort, version 9.0, is harmonized with the hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia cohort. Version 9.2 of the pneumonia mortality measure cohort is, however, not harmonized with the pneumonia payment measure cohort. There is intention to harmonize the pneumonia mortality and payment measure cohorts in the future. We did not include in our list of related measures any non-outcome (for example, process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure). Lastly, this measure and the NQF Inpatient Pneumonia Mortality (AHRQ) Measure #0231 are complementary rather than competing measures. Although they both assess mortality for patients admitted to acute care hospitals with a principal discharge diagnosis of pneumonia, the specified outcomes are different. This measure assesses 30-day mortality while #0231 assesses inpatient mortality. Assessment of 30-day and inpatient mortality outcomes have distinct advantages and uses which make them complementary as opposed to competing. For example the 30-day period provides a broader perspective on hospital care and utilizes standard time period to examine hospital performance to avoid bias by differences in length of stay among hospitals. However, in some settings it may not be feasible to capture post-discharge mortality making the inpatient measure more useable. We have previously consulted with AHRQ to examine harmonization of complementary measures of mortality for patients with AMI and stroke. We have found that the measures are harmonized to the extent possible given that small differences in cohort inclusion and exclusion criteria are warranted on the basis of the use of different outcomes. However, this current measure has been modified from the last endorsed version to include patients with a principal discharge diagnosis of sepsis and a secondary discharge diagnosis of pneumonia that is present on admission. The cohort was also expanded to include patients with a principal discharge diagnosis of aspiration pneumonia. Thus the current measure cohort is no longer harmonized with measure #0231.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0468

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

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

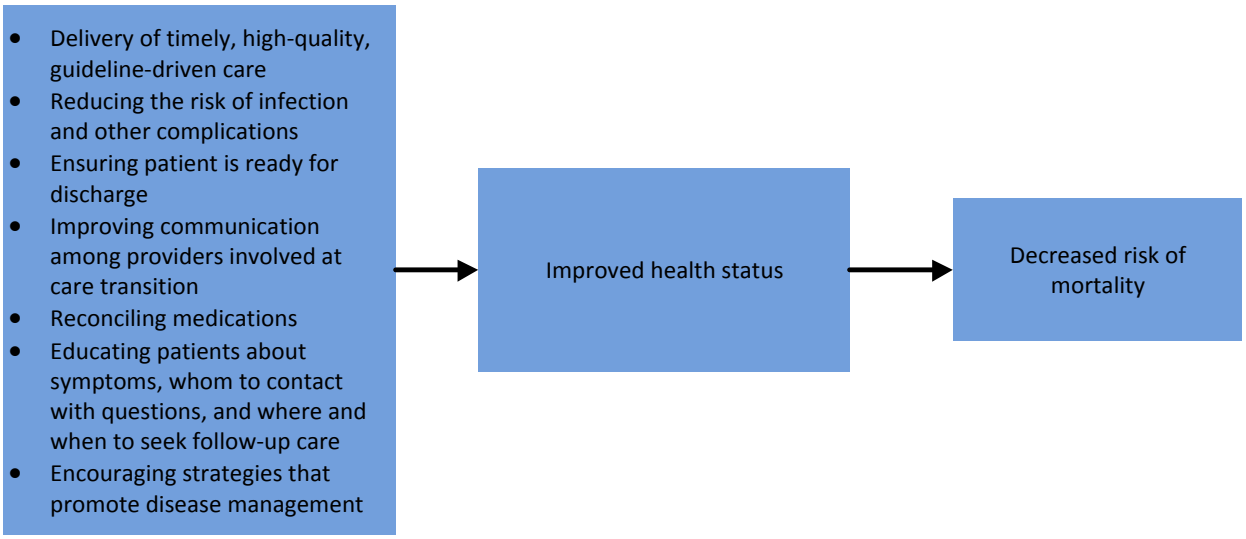
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized mortality 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 mortality 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

In 2010, pneumonia resulted in more than 1 million hospitalizations each year in the United States (Lindenauer et al., 2012; FastStats: pneumonia, CDC). The annual mortality rate (deaths per 100,000 population) is 16.9 (FastStats: pneumonia, CDC). Among patients 65 years [of age] or older in the United States, pneumonia is the second leading cause of hospitalization, and is the leading infectious cause of death (Fry et al., 2005; Bratzler et al., 2011 [1]).

Many current hospital interventions have been shown to be associated with lower risk of death within 30 days of hospital admission (Lee et al., 2014). Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. Measurement of patient outcomes allows for a comprehensive view of quality of care that reflects complex aspects of care such as communication between providers and coordinated transitions to the outpatient environment. These aspects are critical to patient outcomes, and are broader than what can be captured by individual process-of-care measures.

The pneumonia risk standardized mortality rate (RSMR) measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Bratzler et al., 2007 [2]).

The diagram above indicates some of the many care processes that can influence mortality risk. Numerous studies have demonstrated that appropriate (guideline recommended care) and timely treatment for pneumonia patients can reduce the risk of mortality within 30 days of hospital admission (Gleason et al., 1999; Houck et al., 2001; Jha et al., 2007; Lee et al., 2014). Evidence that hospitals have been able to reduce mortality rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect mortality rates (Lee et al., 2014).

References:

Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS One. 2011; 6(4):e17401. [1]

Bratzler DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 2007; 20(2):182-189. [2]

Centers for Disease Control and Prevention. FastStats: pneumonia. Available at: http://www.cdc.gov/nchs/fastats/pneumonia.htm. Accessed August 13, 2015.

Fry AM, Shay DK, Holman RC, et al. Trends in hospitalizations for pneumonia among persons aged 65 years or older in the United States, 1988–2002. JAMA. 2005; 294:2712–2719.

Gleason PP, Meehan TP, Fine JM, et al. Associations between initial antimicrobial therapy and medical outcomes for hospitalized elderly patients with pneumonia. Arch Intern Med. 1999; 159(21):2562-2572.

Houck PM, MacLehose RF, Niederman MS, Lowery JK. Empiric antibiotic therapy and mortality among Medicare pneumonia inpatients in 10 western states: 1993, 1995, and 1997. Chest. 2001; 119(5):1420-1426.

Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. Health Aff (Millwood) 2007 Jul-Aug; 26(4):1104-10.

Lee JS, Nsa W, Hausmann LR, et al. Quality of care for elderly patients hospitalized for pneumonia in the United States, 2006 to 2010. JAMA Intern Med. 2014; 174(11):1806-1814.

Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. JAMA American Medical Association. 2012; 307(13):1405-1413.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency

of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PN_Mortality_Measure_Evidence_Form_v1.0.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality 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 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.

Pneumonia mortality is a priority area for outcomes measure development as it is an outcome that is in part attributable to care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices, as well as increase transparency for consumers.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Distribution of Hospital Pneumonia RSMRs over Different Time Periods
Results for each data year
Characteristic//07/2011-06/20 12//07/2012-06/2013//07/2013-06/2014//07-2011-06/2014
Number of Hospitals// 4,614 // 4,605 // 4,566 // 4,694
Number of Admissions// 460,836 // 482,891 // 434,262 // 1,377,989
Mean (SD)// 16.8 (1.6) // 16.7 (1.6) // 15.5 (1.4) // 16.4 (2.0)
Range (min. – max.)// 11.0-23.9 // 11.1-25.7 // 10.5-22.8 // 8.7-25.4

Minimum// 11.0 // 11.1 // 10.5 // 8.7
10th percentile// 15.0 // 14.8 // 13.9 // 14.0
20th percentile// 15.6 // 15.5 // 14.5 // 14.8
30th percentile// 16.0 // 15.9 // 14.8 // 15.4
40th percentile// 16.3 // 16.3 // 15.1 // 15.8
50th percentile// 16.7 // 16.6 // 15.3 // 16.2
60th percentile// 17.0 // 16.9 // 15.7 // 16.7
70th percentile// 17.4 // 17.3 // 16.0 // 17.3
80th percentile// 17.9 // 17.8 // 16.5 // 17.9
90th percentile// 18.8 // 18.7 // 17.3 // 18.9
Maximum// 23.9 // 25.7 // 22.8 // 25.4

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Distribution of Pneumonia RSMRs by Proportion of Dual Eligible Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (=11.1%) Dual Eligible patients//Hospitals with a high proportion (=24.6%) Dual Eligible patients
Number of Measures Hospitals// 1,081 // 1,083
Number of Patients// 376,625 patients in low-proportion hospitals// 249,621 in high-proportion hospitals
Maximum// 24.6 // 25.4
90th percentile// 18.6 // 19.3
75th percentile// 17.4 // 17.8
Median (50th percentile)// 16.0 // 16.3
25th percentile// 14.8 // 15.0
10th percentile// 13.8 // 13.8
Minimum // 10.7 // 8.7

Distribution of RSMRs by Proportion of African-American Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low Proportion (=0.0%) African-American patients//Hospitals with a high proportion (=8.2%) African-American patients
Number of Measures Entities (Hospitals)// 1,252 // 1,082
Number of Patients//145,085 patients in low-proportion hospitals// 399,018 in high-proportion hospitals
Maximum// 24.5 // 23.4
90th percentile// 18.6 // 19.2
75th percentile// 17.4 // 17.7
Median (50%)// 16.2 // 16.3
25th percentile// 15.1 // 14.9
10th percentile// 14.2 // 13.8
Minimum// 10.9 // 8.7

Distribution of Pneumonia RSMRs by Proportion of Patients with AHRQ SES Index Scores Equal to or Below 45.9:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims and The American Community Survey (2008-2012) data

Characteristic//Hospitals with low proportion of patients with AHRQ SES index score equal to or below 45.9 (=3.9%)//Hospitals with high proportion of patients with AHRQ SES index score equal to or below 45.9 (=53.4%)
Number of Measures Hospitals// 1,082 // 1,082
Number of Patients// 347,875 patients in hospitals with low proportion of patients with AHRQ SES index score equal to or below 45.9 // 218,219 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 45.9
Maximum// 22.7 // 25.4
90th percentile// 18.5 // 19.5
75th percentile// 17.4 // 17.9
Median (50th percentile)// 16.2 // 16.6
25th percentile// 15.0 // 15.2
10th percentile// 13.9 // 14.2
Minimum // 10.7 // 10.4

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
More than one million hospitalizations each year in the United States are caused by pneumonia (Lindenauer et al., 2012; FastStats: pneumonia, CDC). The annual mortality rate (deaths per 100,000 population) is 16.9 (FastStats: pneumonia, CDC). Among patients aged 65 years or over in the United States, pneumonia is the second leading cause of hospitalization, and is the leading infectious cause of death (Fry et al., 2005; Bratzler et al., 2011 [1]). Many current hospital interventions have been shown to be associated with lower risk of death within 30 days of hospital admission (Lee et al., 2014). Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Bratzler et al., 2007 [2]).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS One. 2011; 6(4):e17401. [1]

Bratzler DW, Nsa W, Houck PM. Performance measures for pneumonia: are they valuable, and are process measures adequate. Current Opinion in Infectious Diseases. 2007; 20(2):182-189. [2]

Centers for Disease Control and Prevention. FastStats: pneumonia. Available at: http://www.cdc.gov/nchs/fastats/pneumonia.htm. Accessed August 13, 2015.

Fry AM, Shay DK, Holman RC, et al. Trends in hospitalizations for pneumonia among persons aged 65 years or older in the United States, 1988–2002. JAMA. 2005; 294:2712–2719.

Lee JS, Nsa W, Hausmann LR, et al. Quality of care for elderly patients hospitalized for pneumonia in the United States, 2006 to 2010. JAMA Intern Med. 2014; 174(11):1806-1814.

Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. JAMA American Medical Association. 2012; 307(13):1405-1413.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A. This measure is not a PRO-PM.

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care : Pneumonia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_0468_S2b_Mortality_Data_Dictionary_v0.5_forCMS-635856833973209589.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Annual Updates
1. Updated CC map.

a. Rationale: The ICD-9-CM CC map is updated annually to capture all relevant comorbidities coded in patient administrative claims data.

Planned Update for 2016 public reporting – (changes reflected in this application)
1. Expanded pneumonia cohort to include patients with a principal discharge diagnosis of aspiration pneumonia and those with a principal discharge diagnosis of sepsis (not including severe sepsis) who have a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis.

a. Rationale: The cohort was expanded to capture a broader population of patients admitted for pneumonia and to capture a consistent clinical cohort across hospitals. The cohort expansion responds to changing coding patterns in which patients with pneumonia are increasingly given a principal discharge diagnosis of sepsis. As hospitals increasingly use a principal discharge diagnosis code of sepsis in combination with a secondary discharge diagnosis of pneumonia that is POA, such patients would be excluded from the measure without the cohort expansion. Furthermore, variation in the use of sepsis coding across hospitals could

lead to differential exclusion of pneumonia patients from the measures across hospitals which could bias efforts to comparatively assess hospital quality. (Please see updated 2015 Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia Mortality, version 9.2 for more details on the modifications made to this measure and final measure specifications. The report is posted on the CMS.gov website at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. The pneumonia mortality report (version 9.2) can be found in the AMI, HF, PN, COPD, and Stroke Mortality Update zip folder.)

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator time window: We define the time period for death from any cause within 30 days from the date of admission for the index pneumonia hospitalization.

Denominator time window: This original measure was developed with 12 months of data. The re-specified measure with the expanded pneumonia cohort (version 9.2) was tested with three years of data. The time window can be specified from one to three years. Currently, the measure is publicly reported with three years of index hospitalizations.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The measure counts deaths for any cause within 30 days of the date of admission of the index pneumonia hospitalization.

Identifying deaths in the FFS measure
As currently reported, we identify deaths for FFS Medicare patients 65 years or over in the Medicare Enrollment Database (EDB).

Identifying deaths in the all-payer measure
For the purposes of development of an all-payer measure, deaths were identified using the California vital statistics data file. Nationally, post-discharge deaths can be identified using an external source of vital status, such as the Social Security Administration's Death Master File (DMF) or the Centers for Disease Control and Prevention's National Death Index (NDI).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or over or (2) patients aged 18 years or older. We have specifically tested the measure in both age groups.

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA but no secondary discharge diagnosis of severe sepsis; and with a complete claims history for the 12 months prior to admission. The measure will be publicly reported by CMS for those patients 65 years or older who are Medicare FFS beneficiaries admitted to non-federal hospitals or patients admitted to VA hospitals.

Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:

1. Principal discharge diagnosis of pneumonia, including aspiration pneumonia; or
Principal discharge diagnosis of sepsis (not including severe sepsis), with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA but no secondary discharge diagnosis of severe sepsis.
2. Enrolled in Medicare fee-for-service (FFS)
3. Aged 65 or over
4. Not transferred from another acute care facility
5. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission, and enrolled in Part A during the index admission.

This measure can also be used for an all-payer population aged 18 years and older. We have explicitly tested the measure in both patients aged 18 years and older, and those aged 65 years or over (see Testing Attachment for details).

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:

ICD-9 codes that define patients with pneumonia:

480.0 Pneumonia due to adenovirus
480.1 Pneumonia due to respiratory syncytial virus
480.2 Pneumonia due to parainfluenza virus
480.3 Pneumonia due to SARS-associated coronavirus
480.8 Pneumonia due to other virus not elsewhere classified
480.9 Viral pneumonia, unspecified
481 Pneumococcal pneumonia
482.0 Pneumonia due to Klebsiella pneumoniae
482.1 Pneumonia due to Pseudomonas
482.2 Pneumonia due to Hemophilus influenzae
482.30 Pneumonia due to Streptococcus, unspecified
482.31 Pneumonia due to Streptococcus, group A
482.32 Pneumonia due to Streptococcus, group B
482.39 Pneumonia due to other Streptococcus
482.40 Pneumonia due to Staphylococcus, unspecified
482.41 Methicillin susceptible pneumonia due to Staphylococcus aureus
482.42 Methicillin resistant pneumonia due to Staphylococcus aureus
482.49 Other Staphylococcus pneumonia
482.81 Pneumonia due to anaerobes
482.82 Pneumonia due to escherichia coli
482.83 Pneumonia due to other gram-negative bacteria
482.84 Pneumonia due to Legionnaires´ disease
482.89 Pneumonia due to other specified bacteria
482.9 Bacterial pneumonia, unspecified
483.0 Pneumonia due to mycoplasma pneumoniae
483.1 Pneumonia due to chlamydia
483.8 Pneumonia due to other specified organism
485 Bronchopneumonia, organism unspecified
486 Pneumonia, organism unspecified
487.0 Influenza with pneumonia
488.11 Influenza due to identified 2009 H1N1 influenza virus with pneumonia

ICD-9 codes that define patients with aspiration pneumonia:

507.0 Pneumonitis due to inhalation of food or vomitus

ICD-9 codes that define patients with sepsis (not including severe sepsis [995.92 or 785.52]) (Cohort requires principal discharge diagnosis of sepsis combined with a secondary discharge diagnosis of pneumonia or aspiration pneumonia coded as POA but no

secondary discharge diagnosis of severe sepsis):
038.0 Streptococcal septicemia
038.10 Staphylococcal septicemia, unspecified
038.11 Methicillin susceptible Staphylococcus aureus septicemia
038.12 Methicillin resistant Staphylococcus aureus septicemia
038.19 Other staphylococcal septicemia
038.2 Pneumococcal septicemia [Streptococcus pneumoniae septicemia]
038.3 Septicemia due to anaerobes
038.40 Septicemia due to gram-negative organism, unspecified
038.41 Septicemia due to hemophilus influenzae [H. influenzae]
038.42 Septicemia due to escherichia coli [E. coli]
038.43 Septicemia due to pseudomonas
038.44 Septicemia due to serratia
038.49 Other septicemia due to gram-negative organisms
038.8 Other specified septicemias
038.9 Unspecified septicemia
995.91 Sepsis

ICD-10 codes that define patients with pneumonia:
J12.0 Adenoviral pneumonia
J12.1 Respiratory syncytial virus pneumonia
J12.2 Parainfluenza virus pneumonia
J12.81 Pneumonia due to SARS-associated coronavirus
J12.89 Other viral pneumonia
J12.9 Viral pneumonia, unspecified
J13 Pneumonia due to Streptococcus pneumoniae
J18.1 Lobar pneumonia, unspecified organism
J15.0 Pneumonia due to Klebsiella pneumoniae
J15.1 Pneumonia due to Pseudomonas
J14 Pneumonia due to Hemophilus influenzae
J15.4 Pneumonia due to other streptococci
J15.3 Pneumonia due to streptococcus, group B
J15.20 Pneumonia due to staphylococcus, unspecified
J15.211 Pneumonia due to Methicillin susceptible staphylococcus
J15.212 Pneumonia due to Methicillin resistant staphylococcus
J15.29 Pneumonia due to other staphylococcus
J15.8 Pneumonia due to other specified bacteria
J15.5 Pneumonia due to Escherichia coli
J15.6 Pneumonia due to other aerobic Gram-negative bacteria
A48.1 Legionnaires´ disease
J15.8 Pneumonia due to other specified bacteria
J15.9 Unspecified bacterial pneumonia
J15.7 Pneumonia due to Mycoplasma pneumoniae
J16.0 Chlamydial pneumonia
J16.8 Pneumonia due to other specified infectious organisms
J18.0 Bronchopneumonia, unspecified organism
J18.9 Pneumonia, unspecified organism
J11.00 Influenza due to unidentified influenza virus with unspecified type of pneumonia
J12.9 Viral pneumonia, unspecified
J10.08 Influenza due to other identified influenza virus

ICD-10 codes that define patients with aspiration pneumonia:
J69.0 Pneumonitis due to inhalation of food and vomit

ICD-10 codes that define patients with sepsis (not including severe sepsis [ICD-9 995.92 or 785.52]) (Cohort requires principal discharge diagnosis of sepsis combined with a secondary discharge diagnosis of pneumonia or aspiration pneumonia coded as POA

but no secondary discharge diagnosis of severe sepsis):
A40.9 Streptococcal sepsis, unspecified
A41.2 Sepsis due to unspecified staphylococcus
A41.01 Sepsis due to Methicillin susceptible Staphylococcus
A41.02 Sepsis due to Methicillin resistant Staphylococcus
A41.1 Sepsis due to other specified staphylococcus
A40.3 Sepsis due to Streptococcus pneumoniae
A41.4 Sepsis due to anaerobes
A41.50 Gram-negative sepsis, unspecified
A41.3 Sepsis due to Hemophilus influenzae
A41.51 Sepsis due to Escherichia coli [E. coli]
A41.52 Sepsis due to Pseudomonas
A41.53 Sepsis due to Serratia
A41.59 Other Gram-negative sepsis
A41.89 Other specified sepsis
A41.9 Sepsis, unspecified organism

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The mortality measures exclude index admissions for patients:

1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility;
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
4. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1. The discharge disposition indicator is used to identify patients alive at discharge. Transfers 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. Patient length of stay and condition is identified from the admission claim.

2. Inconsistent vital status or unreliable data are identified if any of the following conditions are met 1) the patient's age is greater than 115 years; 2) if the discharge date for a hospitalization is before the admission date; 3) if the patient has a sex other than 'male' or 'female'.

3. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice enrollment data.

4. Discharges against medical advice (AMA) are identified using the discharge disposition indicator.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Also, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

N/A

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
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 measure employs a hierarchical logistic regression model to create a hospital-level 30-day RSMR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of mortality within 30 days of admission for age, sex, and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of death at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables:
Candidate variables were patient-level risk-adjustors that were expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including age, sex, and indicators of comorbidity and disease severity. For each patient, covariates are obtained from claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using both inpatient and outpatient Medicare FFS claims data. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). 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 index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission.

The final set of risk adjustment variables is:

Demographics
Male
Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts.

Comorbidities
History of Percutaneous Transluminal Coronary Angioplasty (PTCA) (ICD-9 codes V45.82, 00.66, 36.06, 36.07)
History of Coronary Artery Bypass Graft (CABG) (ICD-9 codes V45.81, 36.10–36.16)
Congestive heart failure (CC 80)
Acute myocardial infarction (CC 81)
Other acute/subacute forms of ischemic heart disease (CC 82)
Coronary atherosclerosis or angina (CC 83-84)
Cardio-respiratory failure or shock (CC 78-79)
Hypertension (CC 89, 91)
Stroke (CC 95-96)
Cerebrovascular disease (CC 97-99, 103)
Renal failure (CC 131)
Chronic obstructive pulmonary disease (COPD) (CC 108)
Pneumonia (CC 111-114)
Protein-calorie malnutrition (CC 21)

Dementia or other specified brain disorders (CC 49-50)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Vascular disease and complications (CC 104-105)
Metastatic cancer, acute leukemia and other severe cancers (CC 7-8)
Trauma in last year (CC 154-156, 158-162)
Major psychiatric disorders (CC 54-56)
Chronic liver disease (CC 25-27)
Severe hematological disorders (CC 44)
Iron deficiency or other unspecified anemias and blood disease (CC 47)
Depression (CC 58)
Parkinson's or Huntington's diseases (CC 73)
Seizure disorders and convulsions (CC 74)
Fibrosis of lung or other chronic lung disorders (CC 109)
Asthma (CC 110)
Vertebral fractures (CC 157)
Septicemia/sepsis (CC 2)
Respirator dependence/tracheostomy (CC 77)
Disorders of fluid/electrolyte/acid-base (CC 23)
Delirium and encephalopathy (CC 48)
Decubitus ulcer of skin (CC 148)

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.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 measure estimates hospital-level 30-day all-cause RSMRs following hospitalization for pneumonia using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of index admission using age, sex, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the

underlying risk of a mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution 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 RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on 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 mortality rates or better quality, and a higher ratio indicates higher-than-expected mortality rates or worse quality.

The "predicted" number of deaths (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of mortality. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of deaths (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Krumholz et al., 2005).

References:
Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Models for AMI and HF 30-Day Mortality Methodology. 2005.

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

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. This measure is not based on a sample.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A. This measure is not based on a survey or patient-reported data.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing values are rare among variables used from claims data in this measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Data sources for the Medicare FFS measure:

1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician 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 data source 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 et al., 1992).

3. The American Community Survey (2008-2012): The American Community Survey data is collected annually and an aggregated 5-years data was used to calculate the AHRQ SES composite index score.

4. Data sources for the all-payer update:
For our analyses to examine use in all-payer data, we used all-payer data from California in addition to CMS data for Medicare FFS patients aged 65 years or over (65+) 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 2009, there were 3,193,904 adult discharges from 446 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 pneumonia mortality measure can be applied to all adult patients, including not only FFS Medicare patients aged 65 or over, but also non-FFS Medicare patients aged 18-64 years at the time of admission.

Reference:
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.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A. This measure is not a composite performance measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PN_Mortality_Measure_Testing_Form_v1.0.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0468
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process

☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From:	Measure Tested with Data From:

(*must be consistent with data sources entered in S.23*)	
☐ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☒ other: Census Data/American Community Survey

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The datasets used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). Additionally, census as well as claims data were used to assess socioeconomic factors and race (dual eligible and African American race variables obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score obtained through census data). The dataset used varies by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing?

The dates used vary by testing type; see Section 1.7 for details.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years or over are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of admissions/patients varies by testing type: see Section 1.7 for details

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2)
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or over from the expanded pneumonia cohort (measure version 9.2) and applying the re-specified risk-adjusted model for this group. We then applied the same model for the remaining 50% of patients and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples. In each year of measure reevaluation, we also re-fit the model and compared the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (**Dataset 1** below).

Dataset 1 (expanded cohort, measure version 9.2): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014 (expanded cohort)
Number of Admissions: 1,377,989
Patient Descriptive Characteristics: average age=81.0, % male=46.1
Number of Measured Hospitals: 4,694

First half of split sample
-Number of Admissions: 687,838
-Number of Measured Hospitals: 4,671

Second half of split sample
-Number of Admissions: 690,151
-Number of Measured Hospitals: 4,671

For validity testing (Section 2b2)
Split samples of **Dataset 1**

For testing of measure exclusions (Section 2b3)
Dataset 1

For testing of measure risk adjustment (Section 2b4)
Dataset 1

Dataset 2 (all payer dataset, 2b4.11): California Patient Discharge Data in addition to CMS Medicare FFS data for patients in California hospitals

Dates of Data: January 1, 2009 – December 31, 2009
Number of Admissions: 68,119 (all 18+ total); 25,938 (FFS 65+); 11,033 (non-FFS 65+); 26,593 (all 18-64)
Patient Descriptive Characteristics: mean age=67.9, %male=49.5 (all 18+ total); mean age=80.6, %male=47.9 (FFS 65+); mean age=80.4, %male=48.9 (non-FFS 65+); mean age=48.5, %male=51.1 (all 18-64)
Number of Measured Hospitals: 318 non-Federal acute care hospitals

The measure was applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations. In addition, the unique patient ID number is used to link with state vital statistics records to assess 30-day mortality.

For testing to identify meaningful differences in performance (Section 2b5)
Dataset 1

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.3)
Dataset 1 and **Dataset 3:** The American Community Survey (2008-2012)

We examined disparities in performance according to the proportion of patients in each hospital who were of African-American race and the proportion who were dual eligible for both Medicare and Medicaid insurances. We also used the AHRQ SES index score to study the association between performance measures and socioeconomic status.

Data Elements

- African-American race and dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from CMS enrollment data (**Dataset 1**)
- Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 3**)

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Sociodemographic status incorporates socioeconomic variables as well as race into a more concise term. However, given the fact that socioeconomic risk factors are distinct from race and should be interpreted differently, we have decided to keep "socioeconomic status" and "race" as separate terms.

We selected socioeconomic status (SES) and race variables to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various SES factors and African-American race to worse health status and higher mortality over a lifetime (Adler and Newman, 2002; Mackenbach et al., 2000; Tonne et al., 2005; van Oeffelen et al., 2012). Income, education, and occupational level are the most commonly examined SES variables. However, literature directly examining how different SES factors or race might influence the likelihood of older, insured, Medicare patients of dying within 30 days of an admission for pneumonia is much more limited (Hausmann et al., 2009; Pippins et al., 2007; Polsky et al., 2008). The causal pathways for SES and race variable selection are described below in Section 2b4.3.

The SES and race variables used for analysis were:

- Dual eligible status (**Dataset 1**)
- African-American race (**Dataset 1**)
- AHRQ SES index score using patient 5-digit zip code data (percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room) (**Dataset 3**)

References:

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health affairs (Project Hope).* 2002; 21(2):60-76.

Hausmann LR, Ibrahim SA, Mehrotra A, et al. Racial and ethnic disparities in pneumonia treatment and mortality. Medical care 2009; 47:1009-17.

Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. *European heart journal.* 2000; 21(14):1141-1151.

Pippins JR, Fitzmaurice GM, Haas JS. Hospital characteristics and racial disparities in hospital mortality from common medical conditions. Journal of the National Medical Association 2007; 99:1030-6.

Polsky D, Jha AK, Lave J, et al. Short- and long-term mortality after an acute illness for elderly whites and blacks. Health services research 2008; 43:1388-402.

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. *Circulation.* Jun 14 2005; 111(23):3063-3070.

van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. *European journal of epidemiology.* Aug 2012; 27(8):605-613.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data Element Reliability

In constructing the measure, 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 from logistic regression models across the most recent three years of data (**Dataset 1**).

Measure Score 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 finally comparing the agreement between the two resulting performance measures 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 the 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 (ICC) (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used dataset 1 split sample and calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals using the intra-class correlation as defined by ICC (2, 1) by Shrout and Fleiss (1979).

Using two independent samples provides a stringent 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. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman, 1910; Brown, 1910). We use this to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

References:

Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

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

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; 86:420-428.

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element reliability results (**Dataset 1**)

The frequency of some model variables increased and others decreased between 2011 and 2014, which may reflect an increase or decrease in the rate of specific comorbidities in the FFS population. For example, there was a notable increase in percent frequency for "delirium and encephalopathy" (CC 48) (10.1% to 11.3%), "septicemia/sepsis" (CC2) (10.7%to 11.9%), "cardio-respiratory failure or shock" (CC 78-79) (22.5% to 24.4%), and "renal failure" (CC 131) (30.5% to 31.3%). There was a notable decrease in percent frequency for "congestive heart failure" (CC 80) (39.8% to 38.6%), "coronary atherosclerosis or angina" (CC 83-84) (49.8% to 48.6%), "cerebrovascular disease" (CC 97-99, 103) (24.4% 5o 23.4%), "pneumonia" (CC 111-114) (47.6% to 46.0%), "dementia or other specified brain disorders" (CC 49-50) (38.0% to 36.8%), "severe hematological disorders" (CC 44) (3.6% to 2.2%), and "fibrosis of lung or other chronic lung disorders" (CC 109) (15.3% to 13.1%). Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.

These frequencies are from the expanded cohort and re-specified model (measure version 9.2).

Measure Score Reliability Results (**Dataset 1**)
There were 1,377,989 admissions in the combined 3-year sample, with 687,838 in one randomly selected sample and 690,151 in the other sample. The agreement between the two RSMRs for each hospital was 0.79, which according to the conventional interpretation is "substantial" (Landis and Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data.

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

Lindenauer P, Ross J, Grady J, et al. Reevaluation and Re-Specifications Report Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia: Pneumonia Mortality—Version 9.2; Pneumonia Readmission—Version 8.2, 2015. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed November 20, 2015.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The stability over time of the risk factor frequencies and odds ratios suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates substantial agreement of measure scores across samples using a conservative approach to assessment.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The measure's validity is demonstrated in three manners. The first is clinical and face validity of the cohort expansion. As discussed in the 2015 Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia (Mortality, version 9.2; Readmission, version 8.2) (Lindenauer et al., 2015), made publicly available to support the FY 2016 IPPS rule, the cohort expansion is based on changes in clinical and coding practices that have led to greater numbers of patients with pneumonia being coded with sepsis or aspiration pneumonia as a principal discharge diagnosis. These are patients that the measure is intended to assess, as they fit within the broad clinical category of pneumonia patients and are often treated by the same groups of physicians and staff, using similar treatment strategies.

Moreover, virtually all patients hospitalized with pneumonia meet criteria for sepsis. The expansion was also supported by findings in the literature (Lindenauer et al., 2012; Rothberg et al., 2014).

Second, for a number of claims-based outcome measures, including the original version of this measure, we validated the administrative model with a medical-record based model. In this earlier study, we demonstrated that the rates calculated using the risk adjustment model with claims and medical record data were highly correlated (Krumholz et al., 2006). These analyses, though based on an earlier version of this measure, demonstrated that using comorbidity information from administrative claims data is a valid approach to risk adjustment and specifically, that claims-based risk adjustment adequately assesses the difference in case mix among hospitals. The claims-based measure produced results which were highly correlated with those produced through manual chart audit (Krumholz et al., 2006; Bratzler et al., 2011). The revised pneumonia mortality measure utilizes the same approach as the original (now, currently publicly reported) measure. When developing the expanded cohort for the mortality measure, we re-examined the risk ratios for the risk variables used in the original (or current) measure, which showed that the variables remained predictive of the outcome (that is, mortality). Also, model performance characteristics were similar to those of the current pneumonia mortality measure.

As we demonstrated in our analyses in the 2015 Reevaluation Report (Lindenauer et al., 2015), although the revision is bringing in a large portion of patients currently not included in the measure, the revised version of the measure likely has greater validity in that it has mitigated biases introduced by hospital coding patterns. We confirmed that the approach to risk adjustment was effective, as hospital coding frequency was no longer associated with performance on the revised measure.

Finally, as part of measure validation, we tested the performance of the pneumonia mortality model developed in the first half of a randomly split sample of pneumonia hospitalizations from **Dataset 1** (representing 687,838 admissions from 4,671 hospitals) against the second half of the randomly split sample of pneumonia hospitalizations (representing 690,151 admissions from 4,671 hospitals).

References:

Bratzler D, Normand S, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PloS one. 2011;6(4):e17401.

Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia. 2006. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 19, 2015.

Lindenauer P, Lagu T, Shieh M, Pekow P, Rothberg M. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. Jama. Apr 4 2012; 307(13):1405-1413.

Lindenauer P, Ross J, Strait K, et al. 2015 Reevaluation and Re-Specification Report of the Hospital-Level 30-Day Risk-Standardized Measures Following Hospitalization for Pneumonia Mortality, version 9.2; Pneumonia Readmission, version 8.2. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed November 12, 2015.

Rothberg M, Pekow P, Priya A, Lindenauer P. Variation in diagnostic coding of patients with pneumonia and its association with hospital risk-standardized mortality rates: a cross-sectional analysis. Annals of internal medicine. Mar 18 2014; 160(6):380-388.

ICD-9 to ICD-10 Conversion
Statement of Intent
[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion
ICD-10 codes were identified using 2015 GEM mapping software. We then 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The performance of the first half of the split sample and second half of the split sample from **Dataset 1** was similar. The areas under the receiver operating characteristic (ROC) curve for the two models are 0.724 and 0.727, respectively.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The results between the first half of the split sample and second half of the split sample from **Dataset 1** proved to be similar in each of the model testing that was performed. The ROC results were nearly identical and in line with other mortality models.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In **Dataset 1**:

Exclusion	**N**	**%**	**Distribution across hospitals (N=4,329): Min, 25th, 50th, 75th percentile, max**

1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility	78,836	3.73%	(0.00, 1.92, 3.41, 5.38, 40.00)
2. Inconsistent or unknown vital status or other unreliable data	70	0.00%	(0.00, 0.00, 0.00, 0.09, 1.96)
3. Enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission	39,210	1.85%	(0.00, 0.68, 1.45, 2.54, 19.84)
4. Discharged against medical advice (AMA)	11,505	0.54%	(0.00, 0.00, 0.33, 0.76, 9.21)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusion 1 (patients who were discharged alive on the day of admission or the following day who were not transferred to another acute care facility) accounts for 3.73% of all index admissions excluded from the initial index cohort, the majority of all exclusions, and is meant to ensure a clinically coherent cohort. This exclusion prevents inclusion of patients who likely did not have clinically significant pneumonia.

For **exclusion 2** (patients with inconsistent or unknown vital status or other unreliable demographic [age and gender] data), we do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive. This exclusion accounts for 0.00% of all index admissions excluded from the initial index cohort.

For **exclusion 3** (patients enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission), these patients are likely continuing to seek comfort measures only; thus, mortality is not necessarily an adverse outcome or signal of poor quality care. This exclusion accounts for 1.85% of all index admissions excluded from the initial index cohort.

Exclusions 4 (patients who are discharged AMA) accounts for 0.54% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 36 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Our approach to risk adjustment was 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 measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al., 2007). At the patient level, each model adjusts the log-odds of mortality within 30-days of admission for age, sex, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital-specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. 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.

Clinical Factors
Candidate and Final Risk-adjustment Variables: The original measure was developed using Medicare FFS claims data. Candidate variables were patient-level risk-adjustors that are expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusted for case differences based on the clinical status of the patient at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission. In addition, only comorbidities that conveyed information about the patient at that time or in the 12-months prior, and not complications that arose during the course of the hospitalization were included in the risk-adjustment.

As part of measure reevaluation in 2015, the pneumonia cohort was expanded and the risk model was re-specified. The revised pneumonia mortality measure (version 9.2) will:

1. Retain the clinical comorbidity variables included in the current mortality risk model.

2. Incorporate the following new risk-adjustment variables if present in the 12 months prior to the index admission:
- Septicemia/sepsis (CC2)
- Disorders of fluid/electrolyte/acid-base (CC23)
- Delirium and encephalopathy (CC48)
- Respiratory dependence/tracheostomy (CC77)
- Decubitus ulcer of skin (CC148)

Although none of these risk variables were included in the original measure development and validation, during measure re-evaluation we determined that these risk variables were common (that is, with a prevalence of greater than 10% in the population) and had strong associations with mortality (odds ratio [OR] > 1.5) in the expanded pneumonia cohort. These risk variables also had high levels of face validity in terms of the clinical expectation that these conditions would be associated with worse outcomes if they occurred during the 12 months prior to the index admission.

3. Modify two clinical risk variables as follows:
- add **Pleural effusion/Pneumothorax (CC114)** as part of the pneumonia risk variable (that is, Pleural effusion/Pneumothorax will be added to the currently defined **Pneumonia (CC111 – 113)** risk variable in measure version 9.0**,** which will now be redefined in the risk model as **Pneumonia (CC111 – 114)** in measure version 9.2.)
- add **Respiratory Arrest (CC78)** to the cardio-respiratory failure or shock risk variable (that is, Respiratory Arrest will be added to the currently defined **Cardio-respiratory failure or shock risk variable (CC79)** in measure version 9.0, which will now be redefined in the model as **Cardio-respiratory failure or shock** (**CC78-79)** in measure version 9.2.)

Similar to the rationale for including the CCs noted above, the pneumonia mortality measure will include clinical comorbidity risk variables **CC114 and CC78** because these were common (that is, with a prevalence of greater than 10% in the population) and had strong associations with mortality (OR > 1.5). These risk variables also had high levels of face validity in terms of the clinical expectation that these conditions would be associated with worse outcomes if they occurred during the 12 months prior to the index admission

<u>The final set of risk-adjustment variables is:</u>
• Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts
• Male
• History of PTCA
• History of CABG
• Congestive heart failure
• Acute myocardial infarction
• Other acute/subacute forms of ischemic heart disease
• Coronary atherosclerosis or angina
• Cardio-respiratory failure or shock
• Hypertension
• Stroke
• Cerebrovascular disease
• Renal failure

• Chronic obstructive pulmonary disease (COPD)
• Pneumonia
• Protein-calorie malnutrition
• Dementia or other specified brain disorders
• Hemiplegia, paraplegia, paralysis, functional disability
• Vascular disease and complications
• Metastatic cancer, acute leukemia, and other severe cancers
• Trauma in last year
• Major psychiatric disorders
• Chronic liver disease
• Severe hematological disorders
• Iron deficiency or other unspecified anemias and blood disease
• Depression
• Parkinson's or Huntington's diseases
• Seizure disorders and convulsions
• Fibrosis of lung or other chronic lung disorders
• Asthma
• Vertebral fractures
• Septicemia/sepsis
• Respirator dependence/tracheostomy
• Disorders of fluid/electrolyte/acid-base
• Delirium and encephalopathy
• Decubitus ulcer of skin

Socioeconomic Status (SES) Factors and Race

We selected variables representing socioeconomic status (SES) factors and race for examination based on a review of literature, conceptual pathways, and feasibility. In section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SES and race may influence 30-day mortality.

Our conceptualization of the pathways by which patient SES or race affects 30-day mortality is informed by peer-reviewed literature.

Literature Review of Socioeconomic Status (SES) and Race Variables and Pneumonia Mortality

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization, a literature search was performed with the following exclusion criteria: international studies, articles published more than 10 years ago, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on SES or race and pneumonia mortality. Twenty-seven studies were reviewed by title and abstract, and twenty-one studies were excluded from full-text review. Limited data exist that are specific to SES factors or race pertaining to pneumonia mortality. Studies that have been completed have shown no racial disparity in 30-day pneumonia mortality (Polsky et al., 2008), and equivocal results for in-hospital mortality (Pippins et al., 2007; Hausmann et al., 2009). No studies were identified that evaluated income or other SES variables.

Causal Pathways for Socioeconomic Status (SES) and Race Variable Selection

Although some recent literature evaluates the relationship between patient SES or race and the mortality outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with mortality. The SES factors that have been examined in the mortality literature can be categorized into three domains: (1) patient-level variables, (2)

neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients, and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Alter et al., 2014; Taksler et al., 2012). Neighborhood/community-level variables use information from sources such as the American Community Survey (ACS) as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the Agency for Healthcare Research and Quality (AHRQ) SES index score (Blum et al., 2014). Hospital-level variables measure attributes of the hospital which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital.

The conceptual relationship, or potential causal pathways by which these possible SES risk factors influence the risk of mortality following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider.

1. **Relationship of socioeconomic status (SES) factors or race to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These SES risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job, lack of childcare), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment.

In addition to SES risk factors, studies have shown that worse health status is more prevalent among African-American patients compared with white patients. The association between race and worse health is in part mediated by the association between race and SES risk factors such as poverty or disparate access to care associated with poverty or neighborhood. The association is also mediated through bias in healthcare as well as other facets of society.

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus patients with low income are more likely to be seen in lower quality hospitals, which can contribute to increased risk of mortality following hospitalization (Jha et al., 2011; Reames et al., 2014). Similarly African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner et al., 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to mortality risk is that patients may not receive equivalent care within a facility. For example, African-American patients have been shown to experience differential, lower quality, or discriminatory care within a given facility (Trivedi et al., 2014). Alternatively, patients with SES risk factors such as lower education may require differentiated care – e.g. provision of lower literacy information – that they do not receive.

4. **Influence of socioeconomic status (SES) on mortality risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of mortality without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are complex to distinguish analytically. They also have different implications on the decision to risk adjust or not. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in 1.8, the following SES variables and race were considered:

- African American race (as compared to all others)
- Dual eligible status
- AHRQ SES index score

We assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

References:

Alter DA, Franklin B, Ko DT, et al. Socioeconomic status, functional recovery, and long-term mortality among patients surviving acute myocardial infarction. PloS one 2014; 8:e65130.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation Cardiovascular quality and outcomes 2014; 7:391-7.

Hausmann LR, Ibrahim SA, Mehrotra A, et al. Racial and ethnic disparities in pneumonia treatment and mortality. Medical care 2009; 47:1009-17.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs 2011; 30:1904-11.

Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia. 2006. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 19, 2015.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Pippins JR, Fitzmaurice GM, Haas JS. Hospital characteristics and racial disparities in hospital mortality from common medical conditions. Journal of the National Medical Association 2007; 99:1030-6.

Polsky D, Jha AK, Lave J, et al. Short- and long-term mortality after an acute illness for elderly whites and blacks. Health services research 2008; 43:1388-402.

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

Taksler GB, Keating NL, Cutler DM. Explaining racial differences in prostate cancer mortality. Cancer 2012; 118:4280-9.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Below is a table showing the final variables in the model with associated odds ratios.

Final Model Variables (variables meeting criteria in field 2b4.3) (**Dataset 1**)

Variable	**07/2011-06/2014 OR (95% CI)**
Age minus 65 (years above 65, continuous)	1.05 (1.05, 1.05)
Male	1.24 (1.23, 1.25)
History of Percutaneous Transluminal Coronary Angioplasty (PTCA) (ICD-9 codes V45.82, 00.66, 36.06, 36.07)	0.78 (0.77, 0.80)
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 codes V45.81, 36.10-36.16)	0.90 (0.88, 0.92)
Septicemia/sepsis (CC 2)	0.88 (0.87, 0.90)
Congestive heart failure (CC 80)	1.18 (1.17, 1.19)
Acute myocardial infarction (CC 81)	1.18 (1.15, 1.21)
Other acute/subacute forms of ischemic heart disease (CC 82)	0.96 (0.94, 0.98)
Coronary atherosclerosis or angina (CC 83-84)	0.98 (0.97, 0.99)
Cardio-respiratory failure or shock (CC 78-79)	1.17 (1.15, 1.18)
Hypertension (CC 89, 91)	0.84 (0.83, 0.86)
Stroke (CC 95-96)	1.08 (1.07, 1.10)
Cerebrovascular disease (CC 97-99, 103)	0.94 (0.93, 0.95)
Renal failure (CC 131)	1.05 (1.04, 1.06)
Chronic obstructive pulmonary disease (COPD) (CC 108)	0.93 (0.92, 0.94)
Pneumonia (CC 111-114)	1.11 (1.09, 1.12)
Protein-calorie malnutrition (CC 21)	2.07 (2.05, 2.10)
Dementia or other specified brain disorders (CC 49-50)	1.62 (1.60, 1.64)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.19 (1.17, 1.21)
Vascular disease and complications (CC 104-105)	1.01 (1.00, 1.02)
Metastatic cancer, acute leukemia and other severe cancers (CC 7-8)	2.74 (2.70, 2.78)
Trauma in last year (CC 154-156, 158-162)	1.05 (1.04, 1.06)
Major psychiatric disorders (CC 54-56)	1.08 (1.07, 1.09)
Chronic liver disease (CC 25-27)	1.42 (1.38, 1.47)
Severe hematological disorders (CC 44)	1.19 (1.16, 1.23)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	1.11 (1.10, 1.12)
Depression (CC 58)	0.97 (0.96, 0.98)
Parkinson's or Huntington's diseases (CC 73)	1.21 (1.18, 1.23)
Seizure disorders and convulsions (CC 74)	1.04 (1.02, 1.06)
Fibrosis of lung or other chronic lung disorders (CC 109)	1.06 (1.05, 1.08)
Asthma (CC 110)	0.72 (0.70, 0.73)
Vertebral fractures (CC 157)	1.13 (1.11, 1.15)
Respirator dependence/tracheostomy (CC 77)	0.70 (0.67, 0.73)
Disorders of fluid/electrolyte/acid-base (CC 23)	1.17 (1.16, 1.19)
Delirium and encephalopathy (CC 48)	1.01 (1.00, 1.03)
Decubitus ulcer of skin (CC 148)	1.36 (1.34, 1.38)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Variation in prevalence of the factor across measured entities

The prevalence of SES factors and African-American patients in the pneumonia cohort varies across measured

entities. The median percentage of dual eligible patients is 16.8% (interquartile range [IQR] 11.1%-24.6%). The median percentage of African-American patients is 1.9% (IQR 0.0%-8.2%). The median percentage of patients with an AHRQ SES index score below 45.9 is 16.2% (IQR 3.9%-53.4%).

Empirical association with the outcome (univariate)
The patient-level observed pneumonia mortality rate is higher for dual eligible patients, 17.0%, compared with 16.1% for all other patients. The mortality rate for African-American patients was lower at 15.8% compared with 16.3% for patients of all other races. Similarly the mortality rate for patients with an AHRQ SES index score equal to or below 45.9 was 16.2% compared with 16.3% for patients with an AHRQ SES index score above 45.9.

Incremental effect of SES variables and race in a multivariable model
We then examined the strength and significance of the SES variables and race in the context of a multivariable model. Consistent with the above findings, when we include any of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of each of these variables is small. We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model. Furthermore we find that the addition of any of these variables into the model has little to no effect on hospital performance. We examined the change in hospitals' RSMRs with the addition of any of these variables. The mean absolute change in hospitals' RSMRs when adding a dual eligibility indicator is -0.0001% with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.99996. The mean absolute change in hospitals' RSMRs when adding a race indicator is 0.0000% with a correlation coefficient between RSMRs for each hospital with and without race added of 0.99654. The mean absolute change in hospitals' RSMRs when adding an indicator for a low AHRQ Index of SES score to the model is -0.0022% with a correlation coefficient between RSMRs for hospitals from models with and without an indicator for a low AHRQ Index of SES score is 0.99504.

Overall, we find that among the SES and race variables that could be feasibly incorporated into this model, the relationship between African-American race and patients in the lowest quartile by AHRQ SES index score and mortality is in the opposite direction than what has been the expressed concern of stakeholders regarding adding such adjustment to the models. We also find that the impact of any of these indicators is very small to negligible on model performance and hospital profiling. Given the controversial nature of incorporating such variables into a risk-model we do not support doing so in a case that is unlikely to affect hospital profiling. Given these findings and the complex pathways that could explain any relationship between SES or race with mortality, which do not all support risk-adjustment, we did not incorporate SES or race variables into the measure.

Future reevaluation efforts will explore the relationship between SES and types of pneumonia once ICD-10 data are available.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance (**Dataset 1**)
We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the expanded cohort:

Discrimination Statistics
(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

Calibration Statistics
(3) 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)

We tested the performance of the model for **Dataset 1** described in section 1.7.

References:
Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* **17** (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

For the expanded measure cohort version 9.2 (**Dataset 1**) the results are summarized below:

c-statistic = 0.716

Predictive ability (lowest decile %, highest decile %) = (4.5, 40.3)

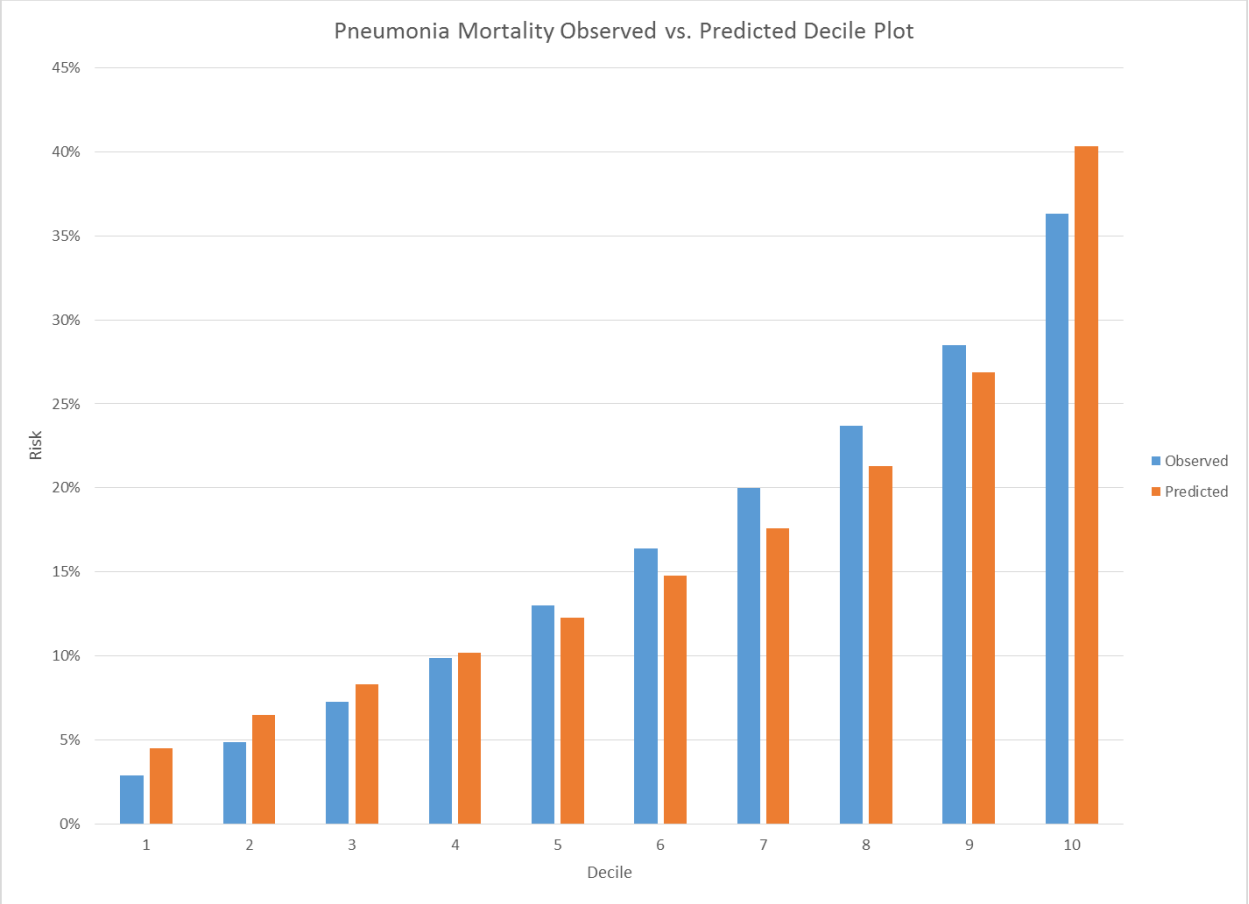
For comparison of model with and without inclusion of SDS factors, see above section.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the expanded cohort (**Dataset 1**) the results are summarized below:
1st half of split sample: Calibration: (0.0457, 0.9526)
2nd half of split sample: Calibration: (0.0496, 0.9504)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2011 to June 2014 (**Dataset 1**).



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e.*, *what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics

The c-statistics of 0.716 indicate good model discrimination (**Dataset 1**). The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Risk Decile Plots

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates good discrimination of the model and fair predictive ability.

Overall Interpretation

Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSMRs accounting for differences in hospital case-mix.

Application to Patients Aged 18 and Older (**Dataset 2**)
We applied the model to all-payer data from California. The analytic sample included 69,247 cases aged 18 and older in the 2009 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.

To help determine whether the measure could be applied to an 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 R2) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.806). 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 good patient-level risk prediction (1.0% to 43.0% vs. 2.0% to 44.0%, respectively, from the bottom decile to the top decile of the prediction values); (b) the c-statistic was nearly identical (0.796 vs. 0.795); and (c) hospital-level risk-standardized rates were highly correlated (r=0.999). 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.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR 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 RSMR 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 cases in the three-year period.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from July 2011-June

2014 (**Dataset 1**), the median hospital RSMR was 16.3%, with a range of 8.7% to 25.4%. The interquartile range was 15.1%-17.5%.

Out of 4,694 hospitals in the U.S., 244 performed "better than the U.S. national rate," 3,814 performed "no different from the U.S. national rate," and 271 performed "worse than the U.S. national rate." 365 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for pneumonia that support measurement to reduce the variation.

Note: The expansion of the cohort to include patients with a principal discharge diagnosis of sepsis who had pneumonia that was present on admission and patients with a principal discharge diagnosis of aspiration pneumonia, resulted in an increase in the mortality rate when compared to the rate from the currently endorsed and publically reported measure, version 9.0 (16.3% and 11.9% respectively). However, over the three years of the measure reporting period, the morality has decreased from 16.7% (July 2011 to June 2012) to 15.4% (July 2013 to June 2014). Despite recent decreases in mortality rates nationally, the mortality rate for the expanded pneumonia cohort measure (version 9.2) is at 16.3%.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note*: This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

N/A

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

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

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements

and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Administrative data are routinely collected as part of the billing process.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees associated with the use of this measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html Payment Program Hospital Value Based Purchasing (HVBP) Program http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting
Program Name, Sponsor: Hospital Inpatient Quality Reporting (Hospital IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital IQR program was originally mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. Initially, the MMA provided for a 0.4 percentage point reduction in the annual market basket (the measure of inflation in costs of goods and services used by hospitals in treating Medicare patients) update for hospitals that did not successfully report. The Deficit Reduction Act of 2005 increased that reduction to 2.0 percentage points.

In addition to giving hospitals a financial incentive to report the quality of their services, the hospital reporting program provides

CMS with data to help consumers make more informed decisions about their health care. Some of the hospital quality of care information gathered through the program is available to consumers on the Hospital Compare website at: www.hospitalcompare.hhs.gov.

Geographic area and number and percentage of accountable entities and patients included:
The Hospital IQR program includes all IPPS non-federal acute care hospitals and VA hospitals in the United States. The number and percentage of accountable hospitals included in the program, as well as the number of patients included in the measure, varies by reporting year. For the data period between 2011 to 2014, the number of hospitals included in the measure with the expanded cohort was 4,694 and the number of admissions was 1,377,989.

Payment Program
Program Name, Sponsor: Hospital Value-Based Purchasing (HVBP) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital Value-Based Purchasing (VBP) Program is a CMS initiative that rewards acute-care hospitals with incentive payments for the quality of care they provide to people with Medicare. It was established by the Affordable Care Act of 2010 (ACA), which added Section 1886(o) to the Social Security Act. The law requires the Secretary of the Department of Health and Human Services (HHS) to establish a value-based purchasing program for inpatient hospitals. To improve quality, the ACA builds on earlier legislation—the 2003 Medicare Prescription Drug, Improvement, and Modernization Act and the 2005 Deficit Reduction Act. These earlier laws established a way for Medicare to pay hospitals for reporting on quality measures, a necessary step in the process of paying for quality rather than quantity.

Geographic area and number and percentage of accountable entities and patients included: More than 3,000 hospitals across the country are eligible to participate in Hospital VBP. The program applies to subsection (d) hospitals located in the 50 states and the District of Columbia and acute-care hospitals in Maryland. Hospital VBP is based on data collected through the Hospital Inpatient Quality Reporting (IQR) Program. More details about the Hospital IQR program are online at https://www.cms.gov/HospitalQualityInits/08_HospitalRHQDAPU.asp.

The following hospitals are excluded from Hospital VBP:
• Hospitals and hospital units excluded from the Inpatient Prospective Payment System, such as psychiatric, rehabilitation, long-term care, children's, and cancer hospitals;
• Hospitals that do not participate in Hospital IQR during the Hospital VBP performance period;
• Hospitals cited by the Secretary of HHS for deficiencies during the performance period that pose an immediate jeopardy to patients' health or safety; and
• Hospitals that do not meet the minimum number of cases, measures, or surveys required by Hospital VBP.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A. This measure is currently publicly reported.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A. This measure is currently publicly reported.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

The median hospital 30-day RSMR for the re-specified pneumonia mortality measure with the expanded cohort (version 9.2) for the 3-year period between July 2011 and June 2014 was 16.3% (IQR 15.1% - 17.5%). The median RSMR decreased by 1.4 absolute percentage points from July 2011- June 2012 (median RSMR: 16.7%) to July 2013- June 2014 (median: RSMR: 15.3%). Hospitals with a high proportion of dual eligible and African American patients achieve a similar range of performance as compared with hospitals with a low proportion of these patients. In addition, hospitals with a low proportion of patients with AHRQ SES index scores below 45.9 perform similarly to hospitals with a high proportion of patients with AHRQ SES index scores below 45.9. These results indicate that both groups of hospitals can perform well on the measure.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure development, model testing, or re-specification. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0231 : Pneumonia Mortality Rate (IQI #20)
0279 : Bacterial Pneumonia Admission Rate (PQI 11)
0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
0708 : Proportion of Patients with Pneumonia that have a Potentially Avoidable Complication (during the episode time window)
2579 : Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on**

interpretability and data collection burden.
The pneumonia mortality measure cohort, version 9.0, is harmonized with the hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia cohort. Version 9.2 of the pneumonia mortality measure cohort is, however, not harmonized with the pneumonia payment measure cohort. There is intention to harmonize the pneumonia mortality and payment measure cohorts in the future. We did not include in our list of related measures any non-outcome (for example, process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure). Lastly, this measure and the NQF Inpatient Pneumonia Mortality (AHRQ) Measure #0231 are complementary rather than competing measures. Although they both assess mortality for patients admitted to acute care hospitals with a principal discharge diagnosis of pneumonia, the specified outcomes are different. This measure assesses 30-day mortality while #0231 assesses inpatient mortality. Assessment of 30-day and inpatient mortality outcomes have distinct advantages and uses which make them complementary as opposed to competing. For example the 30-day period provides a broader perspective on hospital care and utilizes standard time period to examine hospital performance to avoid bias by differences in length of stay among hospitals. However, in some settings it may not be feasible to capture post-discharge mortality making the inpatient measure more useable. We have previously consulted with AHRQ to examine harmonization of complementary measures of mortality for patients with AMI and stroke. We have found that the measures are harmonized to the extent possible given that small differences in cohort inclusion and exclusion criteria are warranted on the basis of the use of different outcomes. However, this current measure has been modified from the last endorsed version to include patients with a principal discharge diagnosis of sepsis and a secondary discharge diagnosis of pneumonia that is present on admission. The cohort was also expanded to include patients with a principal discharge diagnosis of aspiration pneumonia. Thus the current measure cohort is no longer harmonized with measure #0231.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)
Co.4 **Point of Contact:** Karen, Dorsey, karen.dorsey@yale.edu, 203-764-5700-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
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.

Our measure development team consisted of the following members:

Harlan M. Krumholz, M.D., S.M., Principal Investigator
Sharon-Lise T. Normand, Ph.D., M.Sc., Co-Investigator*
Dale W. Bratzler, D.O., M.P.H.**
Jennifer A. Mattera, M.P.H.
Amy S. Rich, M.P.H.
Yongfei Wang, M.Sc., Statistical Analyst
Yun Wang, Ph.D., Senior Biostatistician

*Harvard Medical School, Department of Health Care Policy
**Oklahoma Foundation for Medical Quality

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2006

Ad.3 **Month and Year of most recent revision:** 07, 2015

Ad.4 **What is your frequency for review/update of this measure?** Annual

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.6 **Copyright statement:** N/A

Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** N/A

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0513
Measure Title: Thorax CT—Use of Contrast Material
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: This measure calculates the percentage of thorax computed tomography (CT) studies that are performed with and without contrast out of all thorax CT studies performed (those with contrast, those without contrast and those with both) at each facility. The measure is calculated based on a one-year window of Medicare claims data. The measure has been publicly reported, annually, by the measure steward, the Centers for Medicare & Medicaid Services (CMS), since 2010, as a component of its Hospital Outpatient Quality Reporting (HOQR) Program.
Developer Rationale: This measure will reduce overuse of combination scans of the thorax, as combination scans of the thorax can result in increased exposure to radiation with little clinical benefit. The measure score will guide patient selection of providers, assess quality, and inform quality improvement.

Numerator Statement: The number of thorax CT studies with and without contrast ("combined studies").
Denominator Statement: The number of thorax CT studies performed (with contrast, without contrast, or both with and without contrast) on Medicare beneficiaries within a 12-month time window.
Denominator Exclusions: Indications for measure exclusion include any patients with diagnosis codes associated with: internal injury of chest, abdomen, and pelvis; injury to blood vessels; or crushing injury.

Measure Type: Process
Data Source: Administrative claims
Level of Analysis: Facility, Population : National, Population : State

IF Endorsement Maintenance – Original Endorsement Date: Oct 28, 2008 **Most Recent Endorsement Date:** Jul 31, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

The developer reports the following:

- Updated evidence for this process measure is based on 36 appropriate use criteria (AUC) and two clinical practice guidelines:
  - Thirty-six of the American College of Radiology (ACR) AUC mention the appropriateness of performing CT thorax imaging studies with and without contrast. Thirty-three of the 36 AUC indicate combined CT thorax studies are inappropriate. The quantity, quality, and consistency of the evidence is noted and the recommendations are graded.
  - The National Collaborating Centre for Cancer (NCCC), a center of NICE, recommends patients with known or suspected lung cancer should be imaged using a contrast-enhanced chest CT study to allow for diagnostic confirmation of the disease and to begin staging for cancer treatment. It stated CT chest scans without and with contrast was "Usually Not Appropriate."
  - AIM Specialty Health provides a list of common diagnostic indications for which thorax CT may be appropriately used, including conditions of the chest, pulmonary conditions, mediastinal and hilar conditions, and pleural, chest wall, and diaphragm conditions. No grade assigned.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates: NCCC and AIM guidelines added.

Exception to evidence: NA

Guidance from the Evidence Algorithm: 1→3 →4→5 (eligible for HIGH rating)

Question for the Committee:

- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

- For 2,413 facilities during the 2015 public reporting period, performance rates ranged from 0.0% to 46.5%, with a mean of 3.3%.

Mean Performance (Standard Deviation) 2010-2015

2010	2011	2012	2013	2014	2015
6.9% (13.1)	5.7% (10.5)	6.7% (12.5)	4.8% (8.4)	3.7% (6.7)	3.3% (5.8)

Disparities

- Using 2013 performance data (January 2013 – December 2013), the developer evaluated the effect of patient and facility characteristics on the likelihood of each beneficiary having a combined CT thorax study:
  - The regression model indicates that patient race/ethnicity has a statistically significant effect on the likelihood a patient received an inappropriate combined CT thorax study.
    - African Americans were more likely to undergo inappropriate imaging compared to White beneficiaries (OR 1.234, p=0.000).

        - Latinos were more likely to undergo inappropriate imaging compared to White beneficiaries (OR 1.663, p=0.000).
        - The effect was different for Asians, who were less likely to undergo inappropriate imaging compared to White beneficiaries (OR=0.862, p=0.003).
    - Women were less likely to undergo inappropriate imaging compared to men (OR 0.960, p=0.000).
    - Age also has a statistically significant effect:
        - Compared to beneficiaries aged 60 to 69 years, beneficiaries aged 18 to 29 (OR 0.610, p=0.000), 30 to 39 years (OR 0.741, p=0.000), 70 to 79 years (OR 0.943, p=0.000), 80 to 89 years (OR 0.848, p=0.000), and 90+ years (OR 0.719, p=0.000) were statistically less likely to receive an inappropriate combined CT thorax study.
        - Conversely, patients aged 50 to 59 years were more likely to receive an inappropriate combined CT thorax study (OR 1.087, p=0.000) when compared to beneficiaries aged 60 to 69 years.
        - There was no statistical difference in the likelihood that a patient received an inappropriate combined CT thorax study for patients aged 40 to 49 years compared to beneficiaries aged 60 to 69 years.
    - Facility characteristics also played a role in determining whether a patient received an inappropriate combined CT thorax study.
        - When compared to facilities with fewer than 50 beds (a proxy for facility size), facilities with 51-100 beds (OR 0.889, p=0.000), 101-250 beds (OR 0.805, p=0.000), 251-500 beds (OR 0.780, p=0.000), and 500+ beds (OR 0.528, p=0.000) were less likely to perform inappropriate combined CT thorax studies.
        - A facility's urbanicity impacted a beneficiary's likelihood of having an inappropriate combined CT thorax study – urban facilities were less likely than rural facilities to perform inappropriate combined CT thorax studies (OR 0.576, p=0.000).
        - Teaching (OR 0.732, p=0.000) and major-teaching (OR 0.578, p=0.000) facilities were less likely to perform inappropriate combined CT thorax studies compared to non-teaching facilities.
- The developer chose not to risk adjust or stratify the measure, stating adjusting for these differences would mask underlying differences in quality of care and so is inappropriate. The developer felt there should be no difference in the standard of care for these patients since it was a process measure, and these statistically significant differences are driven by variation in provider practice.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Do you agree with the developer that risk adjustment or stratification is not necessary or appropriate for this measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**This is a process measure focused on reducing the increased exposure to radiation due to unnecessary scans. The evidence provided by the measure developer supports the methodology provided for the measure. Based on the rates reported, it should provide an opportunity to reduce unnecessary scans through patient management.

**OK to not vote

1b. Performance Gap

Comments:

**The gaps described by the measure developer indicated performance gaps based on age, gender, race/ethnicity and rural versus urban settings. It also identified a gap based on the number of beds in the hospital setting. Data on the referenced population subgroups was provided as well as overall data 2010 through 2015. The data did demonstrate over-utilization of thorax scans resulting in disparities in care.

**Yes, gap appears present. Risk adjustment not needed.

1c. High Priority (previously referred to as High Impact)
Comments:
**This is not a composite measure.
**na

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

The developer reported changes to the specifications since the last maintenance review:

- Several exclusion categories were added: The 2013 update to the ACR AUC for blunt abdominal trauma indicated combined studies of the thorax were appropriate for patients diagnosed with blunt abdominal trauma. Consequently, diagnosis codes related to internal injury of chest, abdomen, and pelvis, injury to blood vessels, and crushing injury were proposed as exclusions from the measure. Based on this evidence, the Technical Expert Panel (TEP) recommended excluding patients with a diagnosis of blunt abdominal trauma.
- The numerator for this measure is: *Thorax CT with and without contrast ("combined studies"), occurring on the same day, within a 12-month time window*. The denominator is: *Number of thorax CT studies performed (with contrast, without contrast, or both with and without contrast) on Medicare beneficiaries within a 12-month time window*.
- The CPT, ICD-9, and ICD-10 codes are described in the numerator and denominator details, and included in the value sets excel attachment.
- The calculation algorithm is stated in S.18 and appears straightforward.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Results were not based on a sample, but on 100% of fee for service Medicare claims data. In February-March 2010, CMS conducted a dry run of four outpatient imaging efficiency measures including OP-11: Thorax CT Use of Contrast.
- The dry run included 3.354 hospitals that provided at least hospital-specific reports and 3,060 hospitals that provided at least patient-level data; 3,007 hospitals provided both. During the dry run process, which involved four imaging efficiency measures, 12 of 583 questions/comments (2%) were received on this measure.
- From these findings, and the subsequent low level of inquiries about the technical specification of this measure, CMS inferred that Medicare claims data have provided reliable results about the measure´s numerator and denominator values for the more than 3,000 hospitals participating in Hospital Compare.

Describe any updates to testing:

- The developer conducted new empirical testing at the measure score level.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing: Empirical reliability testing was conducted using a beta-binomial approach that calculated the signal-to-noise ratio. The testing sample was the 3,666 facilities that met minimum case count requirements in 2013.

Results of reliability testing Reliability scores ranged from 30.3% to 100.0%, with a median reliability score of 99.0%, which the developer states is indicative of strong measure reliability. The developer states the results indicate the measure is able to identify true differences in performance between individual facilities.

Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH rating)

Question for the Committee:

o *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity was concluded by the developer.

Describe any updates to validity testing:

- New face validity testing was conducted.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

☐ **Empirical validity testing of the measure score**

Validity testing method: Face validity of the measure score was systematically assessed through survey of a 10-member Technical Expert Panel (TEP). TEP members responded to the following statement:

- The measure helps assess the inappropriate use of thorax CT studies with and without contrast (combined scan). (Response options: *strongly agree*, *agree*, *undecided*, *disagree*, *strongly disagree*, and *do not know*)
- Note, the developer conducted other face validity assessments, but NQF policy permits face validity only at the performance measure score level, as described above.

Validity testing results:

- The developer reports 90% of respondents believe this measure helps assess the inappropriate use of thorax CT studies with and without contrast (combined scans).

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer noted the following:

- The measure excludes patients with diagnosis codes associated with any one of the following
  - Internal Injury of Chest, Abdomen, and Pelvis
  - Injury to Blood Vessels
  - Crushing Injury
- The frequency of exclusions varied across facilities (IQR: 0%-0.54%), however, the prevalence of exclusions for the denominator population was low.
- While median performance does not change significantly, exclusion of patients with a crushing injury or injury to the chest, abdomen, pelvis, or blood vessels is necessary to ensure the face validity of the measure and to align with clinical guidelines. The ACR AUC for Blunt Abdominal Trauma (2012) states that combined CT thorax studies are appropriate for stable patients with blunt abdominal trauma.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer provides the following:

- The developer notes that the performance results from measured entities that perform small numbers of CT thorax studies could have significantly skewed results form only 1-2 "bad" cases.
  - To address this, the developer applies minimum case count requirements before reporting performance scores.
  - The developer states the minimum case count requirements applied for this measure and other imaging efficiency measures "assure a 90% confidence level for the observed rate."
- 190 (5.2%) of 3,666 facilities meeting the minimum case count had performance scores statistically significantly different from the weighted mean (or benchmark). Statistically meaningful difference was defined by the developer as the score was outside the confidence interval (± 1.96 standard deviations). (
- While many facilities have converged around a performance score of between 1% and 4%, more than 5% of facilities continue to have outlying performance; reporting a measure mean (benchmark) provides an opportunity for the outlying facilities to implement efforts to reduced the rate of oversue.

Questions for the Committee:

- *Is the developer's approach to identify the minimum case count and the 90% confidence level appropriate?*
- *Is the method to identify outliers appropriate?*
- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not needed.

2b7. Missing Data

- The analytic files used for measure testing and measure calculation are post-adjudicated claims, and do not have missing data.

Guidance from the Validity Algorithm: 1 → 2 → 3 → 4 → 5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The validity specifications were consistent with the evidence provided. Since the measure developer proposed the updated methodology as an overall measure versus the inclusion of population stratifications, it was reviewed at the level of those individuals that received a scan.

**specs for reliability and validity seem appropriate

2a2. Reliability Testing

Comments:

**Face validity was conducted by the measure developer. Based on a survey of a 10 member TEP, the measure determined a validity of 90%. It is reasonable that the measure results could be used as an indicator of quality of care provided. The validity testing was done at the score level rather than the data element level in accordance with NQF specifications.

**Sufficient validity to measure quality of care in this area.

2b2. Validity Testing

Comments:

**Medicare claims were used so there were no missing data or data elements. Risk adjustment models were not used in calculating the measure.

**appropriate as stated

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:

**Reliability testing was addressed. The measure developer based the reliability on the fact that very few concerns were received related to the reported rates, use of Medicare claims data reduced the error probability, and that the median reliability score was 99%

**performance differences seem obtainable

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer reports the following:

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy
Comments:
**The data elements are routinely collected and reported in claims. Therefore the measure methodology is feasible.
**no concerns

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
- This measure is publically reported in Hospital Outpatient Quality Reporting

Improvement results : The developer reports the median rate of overuse decreased significantly from 2010 to 2015. The median rate of overuse declined by 47.4% (1.9% to 1.0%). From 2010-2015, 2,413 facilities met the minimum case count to be eligible for public reporting in all years; additional facilities met the minimum case count requirements in some, but not all, years. During the 2010 public reporting period, there were more than 76,950 inappropriate CT thorax combined studies performed for Medicare FFS beneficiaries in the United States. This number fell to approximately 39,954 potentially inappropriate CT thorax combined studies in the most recent year of publicly reported data.

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms: The developer did not identify any unintended consequences during measure testing. No evidence of unintended consequences to individuals or populations have been reported to the developer since implementation.

Feedback : No feedback provided on QPS. Measure reviewed by MAP for Hospital Outpatient Quality Reporting Program in 2012 and 2013. The measure was finalized for the program both years. The measure was reviewed for Medicare and Medicaid EHR Incentive Program for Eligible Professionals in 2012, Physician Quality Reporting System (PQRS) in 2013, and Medicare Shared Savings Program in 2015. The measure has not been finalized for these programs.

***Questions for the Committee**:*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments
Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments:
**The measure is in use for hospital reporting by CMS. It was considered for several other reporting requirements such as Meaningful Use. It is a maintenance measure being reviewed. The results of the measure do demonstrate that the reporting may have resulted in a decrease in inappropriate use of scans.
**clearly adequately publicly reported.

Criterion 5: Related and Competing Measures
Related or competing measures • No related or competing measures identified.

Pre-meeting public and member comments

• None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0513
Measure Title: Thorax CT—Use of Contrast Material
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/10/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes
3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
- ☐ Health outcome: Click here to name the health outcome
- ☐Patient-reported outcome (PRO): Click here to name the PRO
  PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors
- ☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Overuse of thorax computed tomography (CT) studies for which a scan with no contrast followed by a scan with contrast is performed.

☐ Structure: Click here to name the structure

☐ Other:

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.
This measure is not a health outcome or PRO performance measure.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).
This measure is not a health outcome or PRO performance measure.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.
The process of identifying thorax CT studies performed concurrently (with a non-contrast study performed first, followed by a study using contrast) is related to improved outcomes, including reduced exposure to radiation, reduced exposure to contrast agents, and more efficient use of imaging resources.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?
☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☒ Other – ***complete section 1a.8***
Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Thirty-six appropriate use criteria (AUC) and two clinical practice guidelines are provided based on their relevance and sample size for the measure. The American College of Radiology (ACR) develops AUC for imaging studies based on clinical condition. As a result, thirty-six of the ACR AUC mention the appropriateness of performing CT thorax imaging studies with and without contrast. In addition to the AUC, the first guideline, developed by the National Collaborating Centre for Cancer (NCCC) (a center of the National Institute for Health and Care Excellence [NICE]), provides practice recommendations for the diagnosis and treatment of non-small-cell (NSCLC) and small-cell lung cancer (SCLC). The second guideline, developed by AIM Specialty Health (a radiology benefit management company), evaluates the appropriate use for imaging of the chest. Citations for the 36 AUC and two guidelines follow:

American College of Radiology AUC:

1 – Francois CJ, Kramer JH, Rybicki RJ, et al. Expert Panel on Vascular Imaging and Interventional Radiology. ACR Appropriateness Criteria abdominal aortic aneurysm: interventional planning and follow-up. [online publication]. Reston (VA): American College of Radiology (ACR). 2012. https://acsearch.acr.org/docs/70548/Narrative/.

2 – Kirsch J, Mohammed TH, Kanne JP, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria acute respiratory illness in immunocompetent patients. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/69446/Narrative/.

3 – Heitkamp DE, Albin MM, Chung JH, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria acute respiratory illness in immunocompromised patients. [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/69447/Narrative/.

4 – Sudakoff GS, Rosen MP, Rybicki RJ, et al. Expert Panels on Vascular Imaging and Gastrointestinal Imaging. ACR Appropriateness Criteria blunt abdominal trauma. [online publication]. Reston (VA): American College of Radiology (ACR). 2012. https://acsearch.acr.org/docs/69409/Narrative/.

5 – Chung JH, Cox CW, Mohammed TH, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria® blunt chest trauma. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/3082590/Narrative/.

6 – Kim H, Rybicki FJ, Majdalany BS, et al. Expert Panel on Vascular Imaging. ACR Appropriateness Criteria blunt chest trauma—suspected aortic injury. [online publication]. Reston (VA): American College of Radiology (ACR). 2014. https://acsearch.acr.org/docs/69410/Narrative/.

7 – Mammen L, Abbara S, Dorbala S, et al. Expert Panel on Cardiac Imaging. ACR Appropriateness Criteria chest pain, suggestive of acute coronary syndrome. [online publication]. Reston (VA): American College of Radiology (ACR). 2014. https://acsearch.acr.org/docs/69403/Narrative/.

8 – Earls JP, White RD, Woodard PK, et al. Expert Panel on Cardiac Imaging. ACR Appropriateness Criteria chronic chest pain—high probability of coronary artery disease. [online publication]. Reston (VA): American College of Radiology (ACR); 2010. https://acsearch.acr.org/docs/69405/Narrative/.

9 – Woodard PK, White RD, Abbara S, et al. Expert Panel on Cardiac Imaging. ACR Appropriateness Criteria chronic chest pain—low to intermediate probability of coronary artery disease. [online publication]. Reston (VA): American College of Radiology (ACR); 2012. https://acsearch.acr.org/docs/69337/Narrative/.

10 – Dyer DS, Mohammed TL, Kirsch J, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria chronic dyspnea—suspected pulmonary origin. [online publication]. Reston (VA): American College of Radiology (ACR); 2012. https://acsearch.acr.org/docs/69448/Narrative/.

11 – Wippold FJ II, Cornelius RS, Aiken AH, et al. Expert Panel on Neurologic Imaging. ACR Appropriateness Criteria cranial neuropathy. [online publication]. Reston (VA): American College of Radiology (ACR). 2012. https://acsearch.acr.org/docs/69509/Narrative/.

12 – Abbara S, Ghoshhajra B, White RD, et al. Expert Panel on Cardiac Imaging. ACR Appropriateness Criteria dyspnea—suspected cardiac origin. [online publication]. Reston (VA): American College of Radiology (ACR); 2010. https://acsearch.acr.org/docs/69407/Narrative/.

13 – Leyendecker JR, Clingan MJ, Eberhardt SC, et al. Expert Panel on Urologic Imaging. ACR Appropriateness Criteria® post-treatment surveillance of bladder cancer [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/69364/Narrative/.

14 – Roberts CC, Kransdorf MJ, Beaman FD et al. Expert Panel on Musculoskeletal imaging. ACR Appropriateness Criteria follow-up of malignant or aggressive musculoskeletal tumors. [online publication]. Reston (VA): American College of Radiology (ACR). 2015. https://acsearch.acr.org/docs/69428/Narrative/.

15 – Ketai LH, Kirsch J, Kanne JP, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria hemoptysis. [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/69449/Narrative/.

16 – Dill KE, George E, Rybicki FJ, et al. Expert Panel on Vascular Imaging and Cardiac Imaging. ACR Appropriateness Criteria® imaging for transcatheter aortic valve replacement. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/3082594/Narrative/.

17 – Moriarty JM, Bandyk DF, Broderick DF, et al. Expert Panels on Vascular Imaging, Neurologic Imaging and Thoracic Imaging. ACR Appropriateness Criteria® imaging in the diagnosis of thoracic outlet syndrome [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/3083061/Narrative/.

18 – Roberts CC, Weissman BN, Appel M, et al. Expert Panel on Musculoskeletal Imaging. ACR Appropriateness Criteria metastatic bone disease. online publication]. Reston (VA): American College of Radiology (ACR). 2012. https://acsearch.acr.org/docs/69431/Narrative/.

19 – Ravenel JG, Mohammed TH, Rosenweig KE, et al. Expert Panel on Thoracic Imaging and Radiation-Oncology-Lung. ACR Appropriateness Criteria non-invasive clinical staging of bronchogenic carcinoma. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/69456/Narrative/.

20 – Mammen L, Woodard PK, Abbara S, et al. Expert Panel on Cardiac Imaging. ACR Appropriateness Criteria® nonischemic myocardial disease with clinical manifestations (ischemic cardiomyopathy already excluded). [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/3082580/Narrative/.

21 – Kalva SP, Rybicki FJ, Dill KE, et al. Expert Panel on Vascular Imaging. ACR Appropriateness Criteria® nontraumatic aortic disease. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/3082597/Narrative/.

22 – Casalino DD, Remer EM, Bishoff JT, et al. Expert Panel on Urologic Imaging. ACR Appropriateness Criteria follow-up of renal cell carcinoma. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/69365/Narrative/.

23 – Lalwani N, Dubinsky T, Javitt MC, et al. Expert Panel on Women's Imaging and Radiation Oncology. ACR Appropriateness Criteria pretreatment evaluation and follow-up of endometrial cancer. [online publication]. Reston (VA): American College of Radiology (ACR). 2013. https://acsearch.acr.org/docs/69459/Narrative/.

24 – Siegel CL, Glanc P, Deshmukh SP, et al. Expert Panel on Women's Imaging and Radiation Oncology-Gynecology. ACR Appropriateness Criteria pretreatment planning of invasive cancer of the cervix. [online publication]. Reston (VA): American College of Radiology (ACR). 2015. https://acsearch.acr.org/docs/69461/Narrative/.

25 – Dewhurst C, Rosen MP, Blake MA, et al. Expert Panel on Gastrointestinal Imaging. ACR Appropriateness Criteria pretreatment staging of colorectal cancer. [online publication]. Reston (VA): American College of Radiology (ACR). 2011. https://acsearch.acr.org/docs/69339/Narrative/.

26 – Leyendecker JR, Clingan MJ, Remer EM, et al. Expert Panel on Urologic Imaging. ACR Appropriateness Criteria® pretreatment staging of invasive bladder cancer. [online publication]. Reston (VA): American College of Radiology (ACR); 2012. https://acsearch.acr.org/docs/69370/Narrative/.

27 – Brown K, Gutierrez AJ, Mohammed TL, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria pulmonary hypertension. [online publication]. Reston (VA): American College of Radiology (ACR). 2012. https://acsearch.acr.org/docs/71095/Narrative/.

28 – Kanne JP, Jensen LE, Mohammed TH, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria radiographically detected solitary pulmonary nodule. [online publication]. Reston (VA): American College of Radiology (ACR); 2012. https://acsearch.acr.org/docs/69455/Narrative/.

29 – Vikram R, Beland MD, Blaufox MD, et al. Expert Panel on Urologic Imaging. ACR Appropriateness Criteria renal cell carcinoma staging. [online publication]. Reston (VA): American College of Radiology (ACR); 2015. https://acsearch.acr.org/docs/69372/Narrative/.

30 – Henry TS, Kirsch J, Kanne JP, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria rib fractures. [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/69450/Narrative/.

31 – Mohammed TH, Kirsch J, Brown K, et al. Expert Panel on Thoracic Imaging. ACR Appropriateness Criteria screening for pulmonary metastases. [online publication]. Reston (VA): American College of Radiology (ACR); 2013. https://acsearch.acr.org/docs/69454/Narrative/.

32 – Moy L, Newell MS, Bailey L, et al. Expert Panel on Breast Imaging. ACR Appropriateness Criteria® stage I breast cancer: initial workup and surveillance for local recurrence and distant metastases in asymptomatic women [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/69496/Narrative/.

33 – Moy L, Newell MS, Bailey L, et al. Expert Panel on Breast Imaging. ACR Appropriateness Criteria stage I breast carcinoma. [online publication]. Reston (VA): American College of Radiology (ACR). 2014. http://www.guideline.gov/content.aspx?id=48278.

34 – Mitchell DG, Javitt MC, Glanc P, et al. Expert Panel on Women's Imaging. ACR Appropriateness Criteria staging and follow-up of ovarian cancer. [online publication]. Reston (VA): American College of Radiology (ACR); 2012. https://acsearch.acr.org/docs/69378/Narrative/.

35 – Oto A, Yacoub JH, Casalino DD, et al. Expert Panel on Urologic Imaging. ACR Appropriateness Criteria staging of testicular malignancy. [online publication]. Reston (VA): American College of Radiology (ACR); 2012. https://acsearch.acr.org/docs/69375/Narrative/.

36 – Hsu JY, Malik SB, Abbara S, et al. Expert Panel on Cardiac Imaging. ACR Appropriateness Criteria suspected infective endocarditis. [online publication]. Reston (VA): American College of Radiology (ACR); 2014. https://acsearch.acr.org/docs/69408/Narrative/.

Guidelines:

1 – National Collaborating Centre for Cancer. Lung cancer. The diagnosis and treatment of lung cancer. London (UK): National Institute for Health and Clinical Excellence (NICE); 2011 Apr. 42 p. (Clinical guideline; no. 121). Guideline available at http://www.guideline.gov/content.aspx?id=34282.

2 – AIM Specialty Health. Appropriate use criteria: imaging of the chest. Chicago (IL): AIM Specialty Health; 2015 May 5. 27 p. [106 references] Guideline available at http://www.aimspecialtyhealth.com/Resources/ajax/get_chest_imaging_guidelines.php.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

ACR AUC:

Thirty-three of the 36 AUC indicate that combined CT thorax studies are inappropriate. *AUC 4* recommends that a combined CT thorax study is appropriate for stable patients with blunt abdominal trauma. *AUC 21* notes that combined CT thorax studies are likely appropriate for patients with nontraumatic aortic disease. Finally, *AUC 31* recommends the use of combined CT thorax studies for screening for pulmonary metastases when the primary malignancy is associated with head and neck carcinoma.

Specific, verbatim AUC ratings are listed below:

ACR Guidelines	Clinical Scenario	With and Without Contrast	With Contrast	Without Contrast
Abdominal Aortic Aneurysm: Interventional Planning and Follow-Up (2012)	Planning for pre-endovascular repair or open repair of abdominal aortic aneurysm.	NR	NR	5
Acute Respiratory Illness in Immunocompetent Patients (2013)	Older than age 40.	1	3	4
	Dementia, any age.	1	3	6
	<40 years, negative physical exam; no signs or risk.	1	1	1
	<40 years, positive physical exam; signs of risk.	1	3	4
	Complicated pneumonia.	2	5	8
	Acute asthma, uncomplicated.	1	1	1
	Acute asthma and suspected pneumonia or pneumothorax.	1	1	2
	Acute exacerbation of chronic obstructive pulmonary disease, "uncomplicated" (no history of coronary artery disease or congestive heart failure, no leukocytosis, fever, or chest pain).	1	1	2

ACR Guidelines	Clinical Scenario	With and Without Contrast	With Contrast	Without Contrast
	Acute exacerbation of chronic obstructive pulmonary disease with one or more of the following: leukocytosis, pain, history of coronary artery disease or congestive heart failure.	1	3	4
Acute Respiratory Illness in Immunocompromised Patients (2014)	Negative, equivocal, or nonspecific chest radiograph.	1	3	9
	Positive chest radiograph, multiple, diffuse, or confluent opacities.	1	3	7
	Positive chest radiograph, noninfectious disease suspected	1	5	8
Blunt Abdominal Trauma (2012)	Unstable patient.	3	3	3
	Stable patient.	7	8	6
	Hematuria >35 RBC/hpf (stable)	7	8	6
Blunt chest trauma (2013)	First-line evaluation. High-energy mechanism.	2	9	5
	Normal anteroposterior (AP) chest radiograph, normal examination, and normal mental status. No high-energy mechanism.	1	5	4
Blunt Chest Trauma—Suspected Aortic Injury (2014)	N/A.	NR	NR	6
Chest Pain, Suggestive of Acute Coronary Syndrome (2014)	N/A.	3	5	2
Chronic Chest Pain—High Probability of Chest Artery Disease (2010)	N/A.	NR	3	NR
Chronic Chest Pain—Low to Intermediate Probability of Coronary Artery Disease (2012)	N/A.	NR	NR	6
Chronic Dyspnea—Suspected Pulmonary Origin (2012)	Any age, nonrevealing or nondiagnostic clinical, standard radiography, and lab studies.	1	5	9
Cranial Neuropathy (2012)	Vocal cord paralysis.	3	6	3
Dyspnea—Suspected Cardiac Origin (2010)	N/A.	NR	5	5
Post-treatment Surveillance of Bladder Cancer (2014)	Superficial transitional cell carcinoma (TCC); no invasion or risk factors.	1	1	1
	Invasive TCC with or without cystectomy.	1	3	3
	Superficial TCC; no invasion with risk factors.	1	1	1
Follow-Up of Malignant or Aggressive Musculoskeletal Tumors (2015)	Lower-risk patient (low grade). Evaluation for metastatic disease to the lung from musculoskeletal primary. Baseline examination at time of diagnosis.	1	1	9
	Lower-risk patient (low grade). Evaluation for metastatic disease to the lung from musculoskeletal primary. Follow-up examination 3–6 months after treatment or surgery.	1	1	9

ACR Guidelines	Clinical Scenario	With and Without Contrast	With Contrast	Without Contrast
	Higher-risk patient (high grade). Evaluation for metastatic disease to the lung from musculoskeletal primary. Baseline examination at time of diagnosis.	1	1	9
Hemoptysis (2014)	Hemoptysis ≥30 cc OR 2 risk factors (>40 years old and >30 pack-year history).	NR	9	6
	Persistent/recurrent hemoptysis (<30 cc) and one risk factor (>40 years old, >30 pack-year history).	NR	9	6
	Massive hemoptysis without cardiopulmonary compromise.	NR	8	5
Imaging for Transcatheter Aortic Valve Replacement (2013)	Pre-intervention planning at the aortic valve plane.	NR	NR	5
Imaging in the Diagnosis of Thoracic Outlet Syndrome (2014)	Imaging in the Diagnosis of Thoracic Outlet Syndrome.	NR	NR	3
Metastatic Bone Disease (2012)	1 cm lung nodule. Non-small-cell carcinoma found at needle biopsy. Now coming for staging and resection.	1	1	1
Non-Invasive Clinical Staging of Bronchogenic Carcinoma (2013)	Non-small-cell lung carcinoma.	1	9	9
	Small-cell lung carcinoma.	1	9	5
Nonischemic Myocardial Disease with Clinical Manifestations (Ischemic Cardiomyopathy Already Excluded) (2013)	Suspected arrhythmogenic cardiomyopathy.	NR	NR	2
	Suspected myocardial infiltrative disease.	NR	NR	1
	Suspected hypertrophic cardiomyopathy.	NR	NR	1
	Suspected acute/subacute myocardial disease.	NR	NR	1
Nontraumatic Aortic Disease (2013)	No variant	7	8	7
Post-treatment Follow- Up of Renal Cell Carcinoma (2013)	Asymptomatic patient; no known metastases.	2	6	5
Pretreatment Evaluation and Follow-Up of Endometrial Cancer (2013)	Newly diagnosed endometrial cancer; when imaging is indicated for treatment planning. (See narrative for clinical scenarios where imaging would be indicated.)	1	8	4
	Post-therapy evaluation in patients with clinically suspected recurrence	1	6	4
Pretreatment Planning of Invasive Cancer of the Cervix (2015)	FIGO stage greater than Ib.	1	6	4
Pretreatment Staging of Colorectal Cancer (2011)	Rectal cancer (small or superficial).	5	8	5
	Rectal cancer: large lesion.	5	8	5
	Colon cancer (other than rectum).	5	8	5
Pretreatment Staging of Invasive Bladder Cancer (2012)	No variant	3	5	5

ACR Guidelines	Clinical Scenario	With and Without Contrast	With Contrast	Without Contrast
Pulmonary Hypertension (2012)	No variant	1	NR	5
Radiographically Detected Solitary Pulmonary Nodule (2012)	Solid nodule equal to or greater than1 cm, low clinical suspicion for cancer.	6	6	8
	Solid nodule equal to or greater than1 cm, moderate to high clinical suspicion for cancer.	6	6	8
	Solid nodule <1 cm, low clinical suspicion for cancer.	5	3	7
	Solid nodule <1 cm, moderate to high clinical suspicion for cancer	5	4	8
Renal Cell Carcinoma Staging (2015)	No variant	3	6	6
Rib Fractures (2014)	Adult. Suspected rib fractures from minor blunt trauma (injury confined to ribs).	1	1	3
	Adult. Suspected rib fractures after CPR.	1	2	5
	Rib pain. Suspected stress fracture.	1	2	3
	Adult. Suspected pathologic rib fracture.	1	2	7
Screening for Pulmonary Metastases (2013)	Primary malignancy: bone and soft tissue sarcoma.	2	5	9
	Primary malignancy: renal cell carcinoma.	1	7	8
	Primary malignancy: testicular cancer.	1	3	7
	Primary malignancy: melanoma.	1	5	8
	Primary malignancy: head and neck carcinoma	9	NR	6
Stage I Breast Cancer: Initial Workup and Surveillance for Local Recurrence and Distant Metastases in Asymptomatic Women (2014)	Newly diagnosed. Initial workup. Rule out metastases.	2	2	2
	Surveillance. Rule out metastases.	1	1	1
Stage I Breast Carcinoma (2014)	Rule out metastases: asymptomatic woman.	2	2	2
	Surveillance. Rule out metastases.	1	1	1
Staging and Follow-Up of Ovarian Cancer (2012)	Pretreatment staging of ovarian cancer (CT chest, abdomen, and pelvis).	3	7	4
	Rule out recurrent ovarian cancer (CT chest, abdomen, and pelvis).	3	7	4
Staging of Testicular Malignancy (2012)	Testis tumor (diagnosed by orchiectomy).	2	7	7
Suspected Infective Endocarditis (2014)	N/A	2	5	1

Guidelines:

Guideline 1 (NCCC) recommends that patients with known or suspected lung cancer should be imaged using a contrast-enhanced chest CT study to allow for diagnostic confirmation of the disease and to begin staging for cancer treatment.

Guideline 2 (AIM Specialty Health) provides a list of common diagnostic indications for which thorax CT may be appropriately used, including conditions of the chest, pulmonary conditions, mediastinal and hilar conditions, and pleural, chest wall, and diaphragm conditions.

Specific, verbatim guideline ratings are listed below:

Guideline 1 (NCCC)
Clinical Condition: Blunt Chest Trauma
Variant 1: First-line evaluation. High-energy mechanism**.**
CT chest without and with contrast: 2 (*Usually Not Appropriate)*
CT chest with contrast: 9 (*Usually Appropriate)*
CT chest without contrast: 5 *(May be Appropriate)*
Variant 2: Normal anteroposterior (AP) chest radiograph, normal examination, and normal mental status. No high-energy mechanism.
CT chest without and with contrast: 1 (*Usually Not Appropriate)*
CT chest with contrast: 5 *(May be Appropriate)*
CT chest without contrast: 4 *(May be Appropriate)*

Guideline 2 (AIM Specialty Health) does not use a rating schema to rank its recommendations.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:
ACR AUC:
Recommendations made within the ACR AUC, were categorized as *Usually Not Appropriate* (numeric scores 1-3), *May be Appropriate* (numeric scores 4-6), *Usually Appropriate* (numeric scores 7-9), or *Not Rated* (indicated as "NR"). For combined CT thorax studies, the ACR AUC assigned 50 of the clinical scenarios a score of *Usually Not Appropriate*.

The following grading scale is used to rate the strength of the conclusions based upon the study design, analysis, and results. The following applies to the recommendations from the AUC:
Rating Scale:
1, 2, 3 Usually not appropriate
4, 5, 6 May be appropriate
7, 8, 9 Usually appropriate

Strength of Evidence:
Category 1 - The study is well-designed and accounts for common biases.
Category 2 - The study is moderately well-designed and accounts for most common biases.
Category 3 - There are important study design limitations.
Category 4 - The study is not useful as primary evidence. The article may not be a clinical study or the study design is invalid, or conclusions are based on expert consensus. For example:
a. The study does not meet the criteria for or is not a hypothesis-based clinical study (e.g., a book chapter or case report or case series description).
b. The study may synthesize and draw conclusions about several studies such as a literature review article or book chapter but is not primary evidence.
c. The study is an expert opinion or consensus document.

Guidelines:

Recommendations made within *Guideline 1* (NCCC) (ranged from a value of 1 (defined as *Usually Not Appropriate)* through 9 (defined as *Usually Appropriate*); specifically, the recommendations made for performing a combined CT thorax study for patients with blunt chest trauma were rated as values of 1 and 2 (defined as *Usually Not Appropriate*). The evidence supporting these recommendations demonstrates consensus within the clinical community that combined CT studies of the thorax (i.e., performing a non-contrast CT followed immediately by a contrast CT) are not appropriate for patients presenting with blunt chest trauma.

The following grading scale is used to rate the strength of the conclusions based upon the study design, analysis and results. The following applies to recommendations from Guideline 1:
Rating Scale:
1, 2, 3 Usually not appropriate
4, 5, 6 May be appropriate
7, 8, 9 Usually appropriate

Strength of Evidence:
Category 1 - The conclusions of the study are valid and strongly supported by study design, analysis and results.
Category 2 - The conclusions of the study are likely valid, but study design does not permit certainty.
Category 3 - The conclusions of the study may be valid but the evidence supporting the conclusions is inconclusive or equivocal.
Category 4 - The conclusions of the study may not be valid because the evidence may not be reliable given the study design or analysis.

Recommendations made within *Guideline 2* (AIM Specialty Health) also demonstrate consensus within the clinical community that there are limited clinical circumstances for which imaging of the chest using computed tomography is appropriate. As previously noted, *Guideline 2* does not assign grades to its recommendations.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)
All relevant information about the grades and associated definitions for the 36 AUC and 4 guidelines provided has been included in **section 1a.4.3**.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**
Citations and URLS are the same as those noted in **section 1a.4.1**.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?
☒ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION
1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):
This measure is not based on a recommendation from the US Preventive Services Task Force.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.
This measure is not based on a recommendation from the US Preventive Services Task Force.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:
This measure is not based on a recommendation from the US Preventive Services Task Force.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)
This measure is not based on a recommendation from the US Preventive Services Task Force.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

This measure is not based on a recommendation from the US Preventive Services Task Force.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):
Guidelines are evidenced based; details are provided in **Section 1a.7**.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**
Guidelines are evidenced based; details are provided in **Section 1a.7**.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?
The clinical focuses of the AUC for which these systematic reviews were performed are diagnoses and conditions affecting or related to the thorax region.

Methodologic Approach for the Systematic Reviews that Support the Cited AUC:
ACR performs a systematic review of the literature for each AUC it publishes; these reviews assess the benefits and harms of the imaging procedure, using scientific evidence, clinical judgment, and expert consensus to develop each recommendation. Literature supporting each appropriateness criteria is obtained from indexed articles in MEDLINE (through PubMed), and is reviewed by both ACR staff and the authors of the appropriateness criteria. Literature-search summaries are made publicly available in the form of evidence tables.

The systematic reviews for *AUC #1-36* are the basis of the responses for this section. The evidence tables for each are available on the ACR website, and can be provided by the measure developer upon request.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:
Grades for the evidence provided *AUC #1-36* can be found in section **1a.4.3**.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.
Grades for the evidence provided from *AUC #1-36* can be found in section **1a.4.3**.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Date ranges vary by AUC

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)
The systematic reviews for each AUC include six categories of study designs: 1) review/other study types focused on diagnosis, 2) review/other study types focused on treatment, 3) observational studies focused on diagnosis, 4) observational studies focused on treatment, 5) experimental studies focused on diagnosis, and 6)

experimental studies focused on treatment. In combination, there were more than 850 reviews and other study types focused on diagnosis, 140 reviews and other study types focused on treatment, 1,110 observational studies focused on diagnosis, 80 observational studies focused on treatment, 30 experimental studies focused on diagnosis, and 10 experimental studies focused on treatment.

As an example, the systematic review for *AUC 7*, focused on chest pain suggestive of acute coronary syndrome, includes eight experimental studies, 27 observational studies, and 27 reviews or other study designs. Experimental studies ranged in size from 105 to 9,212 patients. Observational studies ranged in size from 31 to 736 patients. For reviews and other study designs, the sample size ranged from 97 to 4,278 patients.

The quantity and quality of the body of evidence is further bolstered by the literature used for guideline development for the other guidelines that support this measure.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

See response to **section 1a.7.5**. Results cited in this body of evidence are consistent across studies and guidelines, providing evidence that combined CT thorax studies are inappropriate absent "red flag" conditions.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of %ages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Given the high costs associated with performing unnecessary combined CT thorax studies, the potentially harmful exposure to ionizing radiation, and risk of contrast-induced nephropathy, the overall net benefit in reducing overuse of combined thorax CT studies is a reduction in cost and a reduction in morbidity, per beneficiary.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

No harms in measure implementation were identified to counter the net benefit of the measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

In addition to the two guidelines and 36 AUC cited above, a review of the clinical literature was conducted during the measure contractor's annual review of the literature for additional evidence and/or new studies that substantiate the measure's intent. Citations and summaries for the four items included in this review can be found in **section 1a.8.2**. Some of these four studies have been published since the period of guideline development. Results cited in these studies are consistent across studies and with the guidelines cited above, which further supports the measure's specifications.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

A review of the clinical literature was conducted during the measure contractor's annual review of the literature for additional evidence and/or new studies that support the measure's intent. The measure contractor identified relevant peer-reviewed publications by searching the PubMed MEDLINE database from January 1, 2013 to

January 16, 2015, limiting included results to those published in the English language and that had abstracts available in PubMed. The search initially identified 103 articles; a further review by the contractor's clinical and measure-development team resulted in the inclusion of four articles in the body of evidence below. Citations and summaries for the four items included in this review can be found in **section 1a.8.2**.

1a.8.2. Provide the citation and summary for each piece of evidence.

Kurabayashi M, Okishige K, Ueshima D, et al. Diagnostic utility of unenhanced computed tomography for acute aortic syndrome. Circ J. 2014; 78(8): 1928-1934.

Kurabayashi et al. examined the diagnostic value of unenhanced computed tomography (CT) for diagnosing acute aortic dissection (AAD) and ruptured thoracic aortic aneurysm (TAA). The study team examined 219 consecutive patients who visited the emergency room with suspected acute aortic syndrome (AAS) because of chest or back pain and who underwent both unenhanced and contrast-enhanced 64-row multi-detector CT. The unenhanced CT findings were evaluated by the cardiologist on duty who was blind to the findings of the contrast-enhanced CT. The study ultimately determined that unenhanced CT is a good tool for ruling AAS in, but the false-negative rate of 6.7% is high for ruling AAS out. The study does not make recommendations specific to the use of a combined study.

Levin, D, Rao, V, Parker, L, Frangos, A. Are "double" CT scans of the thorax being overused? J Am Coll Radiol. 2014; 11: 788-790.

Levin et al. performed a claims analysis to determine what proportion of all thoracic CT scans are combined scans in the Medicare population. The study team utilized the Medicare Part B Physician/Supplier Procedure Summary Master Files for 2001 to 2011 to calculate utilization rates per 1,000 beneficiaries for combined scans. The team found that the utilization rate of combined scans increased from 2001 through 2006, remained steady in 2007, but then decreased sharply thereafter. The authors concluded that in 2011, 4.2% of thoracic CT scans nationwide were performed both without and with contrast. The percentage dropped by almost one-third from 2006 to 2011 suggesting that the practice is declining.

Mathias JS, Feinglass J, Baker DW. Variations in US hospital performance on imaging-use measures. Med Care. 2012;50(9):808-14.

Mathias et al. used data from the Hospital Outpatient Quality Reporting (HOQR) Program to assess consistency in performance across measures. The study focused on whether higher imaging use could be associated with certain hospital characteristics. The study examined associations between hospital characteristics and higher use of imaging by drawing on 2008 HOQR data linked with the 2009 American Hospital Association survey. The study found that use of imaging varied widely and was weakly correlated across most measures. Of note, hospitals with low volume (<25th percentile) were more likely to report higher imaging use than hospitals of medium volume (25th to 75th percentile). Of particular interest, rural hospitals were more likely to report highest-decile use on CT thorax, in addition to several other measures. For-profit hospitals were also more likely to report highest-decile use on CT thorax measures. The study concluded that there are significant variations in use of imaging, with some hospitals reporting exceptionally high use.

Yoo SY, Kim Y, Cho HH, et al. Dual-energy CT in the assessment of mediastinal lymph nodes: comparative study of virtual non-contrast and true non-contrast images. Korean J Radiol. 2013; 14(3): 532-539.

Yoo et al. evaluated the reliability of virtual non-contrast (VNC) images reconstructed from contrast-enhanced, dual-energy scans compared with true non-contrast (TNC) images in the assessment of high CT attenuation or calcification of mediastinal lymph nodes. The study concluded that VNC images may be useful in the evaluation of mediastinal lymph nodes by providing additional information of high CT attenuation of nodes, although it is underestimated compared with TNC images.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_0513_Measure_Evidence_Form_2015-12-10.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure will reduce overuse of combination scans of the thorax, as combination scans of the thorax can result in increased exposure to radiation with little clinical benefit. The measure score will guide patient selection of providers, assess quality, and inform quality improvement.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Analysis of Medicare fee-for-service (FFS) claims data indicates variation in the use of combined CT thorax studies. For the 2015 public reporting period, performance rates ranged from 0.0% to 46.5%, with a mean of 3.3%.

The data presented below represent information for the 2,413 facilities whose denominator counts met minimum case count requirements for all years included in the table.

Further details on the descriptive statistics for longitudinal facility performance are included below:

Public Reporting Year	2010	2011	2012	2013	2014	2015
Measurement Period	January 2008 – December 2008		January 2009 – December 2009		January 2010 – December 2010	
	January 2011 – December 2011		July 2012 – June 2013		July 2013 – June 2014	
Facilities	2,413	2,413	2,413	2,413	2,413	2,413
Minimum Value	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1st Percentile	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5th Percentile	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10th Percentile	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
25th Percentile	0.5%	0.4%	0.5%	0.3%	0.2%	0.1%
Median	1.9%	1.7%	1.8%	1.5%	1.2%	1.0%
75th Percentile	6.5%	5.8%	6.4%	5.1%	4.1%	3.5%
90th Percentile	18.3%	16.3%	18.7%	13.8%	9.7%	9.5%
95th Percentile	33.9%	27.7%	32.5%	22.6%	16.7%	15.1%
99th Percentile	69.8%	55.3%	66.0%	41.7%	34.9%	30.1%
Maximum Value	89.3%	90.6%	90.5%	81.3%	64.3%	46.5%
Mean Performance (Standard Deviation)	6.9% (13.1)		5.7% (10.5)		6.7% (12.5)	4.8% (8.4)	3.7% (6.7)	3.3% (5.8)

One of the intentions for reporting this measure is to identify facilities with significant outlying performance. As shown in the table above, many facilities cluster around a value of 0 to 2 percent; however, outlying performance persists, indicating there are facilities for which there is a notable rate of overuse.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of**

measurement.
Data have been included in Section 1b.2; these data represent national performance over time, from 2010 to 2015.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Using 2013 performance data (January 2013 – December 2013), we evaluated the effect of patient and facility characteristics on the likelihood of each beneficiary having a combined CT thorax study. Using a logistic regression model, we assessed the impact of patient and facility characteristics for the 1,667,550 CT thorax studies performed in 2013 and found that patient race/ethnicity, patient gender, patient age, and facility characteristics had a significant relationship with the rate of inappropriate combined CT thorax studies.

The regression model indicates that patient race/ethnicity has a statistically significant effect on the likelihood that a patient received an inappropriate combined CT thorax study. African Americans were more likely to undergo inappropriate imaging compared to White beneficiaries (OR 1.234, p=0.000). Similarly, Latinos were more likely to undergo inappropriate imaging compared to White beneficiaries (OR 1.663, p=0.000). The effect was different for Asians, who were less likely to undergo inappropriate imaging compared to White beneficiaries (OR=0.862, p=0.003).

The regression model also indicates that patient gender is a significant factor. Women were less likely to undergo inappropriate imaging compared to men (OR 0.960, p=0.000).

Patient age also had a statistically significant effect within the regression model. Compared to beneficiaries aged 60 to 69, beneficiaries aged 18 to 29 (OR 0.610, p=0.000), 30 to 39 (OR 0.741, p=0.000), 70 to 79 (OR 0.943, p=0.000), 80 to 89 (OR 0.848, p=0.000), and 90+ (OR 0.719, p=0.000) were statistically less likely to receive an inappropriate combined CT thorax study. Conversely, patients aged 50 to 59 were more likely to receive an inappropriate combined CT thorax study (OR 1.087, p=0.000) when compared to beneficiaries aged 60 to 69. There was no statistical difference in the likelihood that a patient received an inappropriate combined CT thorax study for patients aged 40 to 49 compared to beneficiaries aged 60 to 69.

Facility characteristics also played a role in determining whether a patient received an inappropriate combined CT thorax study. When compared to facilities with fewer than 50 beds (a proxy for facility size), facilities with 51 – 100 beds (OR 0.889, p=0.000), 101 - 250 beds (OR 0.805, p=0.000), 251 – 500 beds (OR 0.780, p=0.000), and 500+ beds (OR 0.528, p=0.000) were less likely to perform inappropriate combined CT thorax studies. Similarly, a facility's urbanicity impacted a beneficiary's likelihood of having an inappropriate combined CT thorax study – urban facilities were less likely than rural facilities to perform inappropriate combined CT thorax studies (OR 0.576, p=0.000). Finally, teaching (OR 0.732, p=0.000) and major-teaching (OR 0.578, p=0.000) facilities were less likely to perform inappropriate combined CT thorax studies compared to non-teaching facilities.

While the regression model identified subpopulations of patients and facilities for which there are statistically significant differences in the rate of inappropriate combined CT thorax studies, these disparities do not indicate a need for adjustment of the measure specifications. Adjusting for these differences would mask underlying differences in quality of care. As this is a process measure, there should be no difference in the standard of care for these patients; we believe these statistically significant differences are driven by variation in provider practice. Consequently, we do not believe risk adjustment or stratification is necessary or appropriate for this measure.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Limited information exists within the literature related to disparities in care for patients undergoing concurrent thorax CT studies. Mathias et al. (2012) used data from the HOQR Program to assess consistency in performance across measures, focusing on whether higher imaging use could be associated with certain hospital characteristics. To do so, the study team examined associations between hospital characteristics and higher use of imaging, drawing on 2008 HOQR data linked with data from the 2009 American Hospital Association survey. Mathias and his team found that use of imaging varied widely and was weakly correlated across most measures. Of note, hospitals with low volume (<25th percentile) were more likely to report higher imaging use than were hospitals of medium volume (25th to 75th percentile). Of particular interest, rural hospitals were more likely to report highest-decile use on combined thorax CT, in addition to several other measures. For-profit hospitals were also more likely to report highest-decile use on thorax CT measures. The study authors concluded that there are significant variations in use of imaging, with some hospitals reporting exceptionally high use.

Despite evidence identified in the literature indicating disparities in care for certain facility types, these disparities do not indicate a need for adjustment of the measure specifications. Adjusting for these differences would mask underlying differences in quality of care. As this is a process measure, there should be no difference in the standard of care for these patients; we believe these statistically significant differences are driven by variation in provider performance. Consequently, we do not believe risk adjustment or stratification is necessary or appropriate for this measure.

REFERENCES

1.) Mathias JS, Feinglass J, Baker DW. Variations in US hospital performance on imaging-use measures. Med Care. 2012;50(9):808-14.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Other

1c.2. **If Other:** Safety

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Published literature indicates that there is significant inappropriate use of CT thorax studies performed with and without contrast. The potentially preventable use of combined thorax CT studies, defined as those that are performed both without and with contrast agents for the evaluation of solid organs and body cavities, represents a serious inefficiency of practice and a patient-safety concern. The evidence base supporting this measure indicates that a thorax CT scan should be performed either without or with contrast, but not both, except in limited clinical circumstances.

A paper presented at the 2011 Radiological Society of North America (RSNA) meeting by Sharpe et al. aimed to determine whether the previously seen rapid growth patterns of advanced imaging in the Medicare program CT, magnetic resonance [MRI], and nuclear medicine [NM]) have changed in recent years. These researchers examined the nationwide Medicare Part B fee-for-service (FFS) databases for 2000 through 2009, aggregating all discretionary codes for CT, MRI, and NM. Global and professional component claims were tabulated. Technical component (TC) claims were excluded to avoid double counting. Rates of use per 1,000 Medicare beneficiaries in all places of service were calculated for each modality each year. Compound annual growth rates (CAGRs) were calculated for the periods 2000 through 2006 and 2007 through 2009, and compared.

As noted in the Sharpe presentation, there was nationwide, rapid growth in CT, MRI, and NM utilization rates per 1,000 Medicare beneficiaries from 2000 through 2006. However, from 2007 through 2009, there was dramatic curtailment of growth in CT and MRI. Composite growth of all three modalities together after 2006 was very modest (CAGR of 1.4 percent). Limiting the analysis to CT, the rate per 1,000 rose from 325 in 2000 to 576 in 2006, rising again to 636 in 2009. This represents annual growth of +10.0 percent from 2000 through 2006, dropping to +3.4 percent from 2007 through 2009.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1.) Sharpe Jr RE, Levin DC, Parker L, Sunshine JH, and Rao VM. Is Growth in Advanced Imaging at an End?, presented at the Radiological Society of North America (RSNA) annual meeting, November 27, 2011. Abstract retrievable at http://archive.rsna.org/2011/11005052.html.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

This measure is not a PRO-PM measure.

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Cancer, Cardiovascular : Congestive Heart Failure, Pulmonary/Critical Care, Pulmonary/Critical Care : Asthma, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Pulmonary/Critical Care : Critical Care, Pulmonary/Critical Care : Dyspnea, Pulmonary/Critical Care : Pneumonia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Overuse, Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_0513_Measure_Value_Sets_2015-12-10.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
As part of the annual measure maintenance and review process, several exclusion categories were added to the measure in 2014. The 2013 update to the American College of Radiology Appropriate Use Criteria for blunt abdominal trauma indicated combined studies of the thorax were appropriate for patients diagnosed with blunt abdominal trauma. Consequently, diagnosis codes related to internal injury of chest, abdomen, and pelvis, injury to blood vessels, and crushing injury were proposed as exclusions from the measure. Based on this evidence, the contractor's Technical Expert Panel (TEP) recommended excluding patients with a diagnosis of blunt abdominal trauma.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of thorax CT studies with and without contrast ("combined studies").

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: Thorax CT with and without contrast ("combined studies"), occurring on the same day, within a 12-month time window.

Denominator: Thorax CT studies performed (with contrast, without contrast, or both with and without contrast) within a 12-month time window.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The numerator is defined by the following CPT Code:

71270- Thorax CT with and without contrast.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The number of thorax CT studies performed (with contrast, without contrast, or both with and without contrast) on Medicare beneficiaries within a 12-month time window.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The denominator is defined by the following CPT codes:

71250- Thorax CT without contrast.
71260- Thorax CT with contrast.
71270- Thorax CT with and without contrast.

Global and TC claims should be considered in order to capture all outpatient volume facility claims, typically paid under the Outpatient Prospective Payment System (OPPS)/Ambulatory Payment Classifications (APC) methodology, and to avoid double counting of professional component claims (i.e., 26 modifier).

A technical unit can be identified by a modifier code of TC. A global unit can be identified by the absence of a TC or 26 modifier code.

Thorax CT studies can be billed separately for the technical and professional components, or billed globally, which includes both the professional and TCs.

Professional component claims will outnumber TC claims due to over-reads.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Indications for measure exclusion include any patients with diagnosis codes associated with: internal injury of chest, abdomen, and pelvis; injury to blood vessels; or crushing injury.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Indications for measure exclusion include any patients with the following diagnosis codes:

Internal Injury of Chest, Abdomen, and Pelvis
ICD-9 Codes: 860-869
Injury to Blood Vessels
ICD-9 Codes: 901-902
Crushing Injury
ICD-9 Codes: 926, 929

Crushing Injury of unspecified hip with thigh
ICD-10 Codes: S77.20*
Injuries to the thorax
ICD-10 codes: S21.301*-S21.459*, S25.00X*-S27.9XX*
Injuries to the abdomen, lower back, lumbar spine, pelvis, and external genitals
ICD-10 codes: S31.001*, S31.021*, S31.031*, S31.041*, S31.051*, S31.600*-S31.659*, S35.00X*-S38.1XX*

For ICD-10 exclusion codes, an appending asterisk (*) represents a wildcard for that digit.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable; this measure does not stratify its results.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable; this measure does not risk adjust.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
No risk model specifications are provided, as risk adjustment or stratification is not necessary for this measure.

S.16. **Type of score:**
Other (specify):
If other: Percentage

S.17. **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

S.18. **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.)*
This measure calculates the percentage of thorax studies that are performed with and without contrast, out of all thorax studies performed (those with contrast, those without contrast, and those with both). The measure is calculated based on a one-year window of hospital outpatient claims data, as follows:
1. Select hospital outpatient claims with a CPT code for any thorax CT study (i.e., 71250- Thorax CT without Contrast, 71260- Thorax CT with Contrast, or 71270- Thorax CT with and without Contrast) on a revenue line item
2. Exclude professional component only claims with modifier = '26'
3. Exclude cases with one or more exclusion diagnoses included on claim
4. Set denominator counter = 1
5. Set numerator counter = 1 if CPT code = 71270 thorax CT studies with and without contrast (combined studies)
6. Aggregate denominator and numerator counts by Medicare provider number
7. Measure = numerator counts/denominator counts [The value should be recorded as a percentage]

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure relies exclusively on 100% Medicare FFS standard analytical file (SAF) data; no sampling of beneficiaries was performed.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This measure does not use survey data.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)

Required for Composites and PRO-PMs.
The measure development contractor does not make any adjustments for missing data. The measure relies on Medicare claims data, which are used for payment purposes for services rendered by a provider. The data undergo prepayment claims analysis and post payment audits, as part of the CMS administrative process. The analytic files used by the measure developer are post-adjudicated claims.

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure was initially constructed using the 100 percent Medicare FFS outpatient SAFs from 2007. These outpatient SAFs contain the claims data on imaging utilization and performed in hospital outpatient departments (including emergency department services), which are necessary to attribute the measure to specific facilities. Public reporting of the measure currently uses the 100 percent Medicare FFS outpatients SAFs from 2013 and 2014.

S.25. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National, Population : State

S.27. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospital/Acute Care Facility, Imaging Facility
If other:

S.28. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable; this is not a composite measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_0513_Measure_Testing_Form_2015-12-10.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0513
Measure Title: Thorax CT—Use of Contrast Material
Date of Submission: 12/10/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**

- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

<u>Note</u>: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
We tested the measure using 2013 Medicare fee-for-service (FFS) data from the 100% Outpatient Standard Analytic File (SAF-O).

Facility Analysis

a. Datasets used to define the initial patient population (denominator):
   - *SAF-O:* CORE and Lewin defined the initial patient population based on the 2013 100% SAF-O file. The initial patient population includes all claims for a CT thorax study from January 1, 2013-December 31, 2013, provided in a hospital outpatient setting. This dataset also includes unique patient and facility identifiers.
   - *Enrollment and denominator files:* This dataset contains Medicare FFS enrollment, demographic, and death information for patients identified in the above file.
   - *Provider of services (POS) file:* The POS file contains data on facility characteristics including urbanicity, bed count, and teaching status.

b. Datasets used to capture the numerator:
   - *SAF-O:* For patients included in the initial patient population, CORE and Lewin identified numerator cases, patients with a combined CT thorax study (performed with and without contrast), by searching the 2013 100% SAF-O file for the CPT code 71270 – Thorax CT With and Without Contrast.

c. Datasets used to identify measure exclusions:
   - *SAF-O:* For patients included in the initial patient population, CORE and Lewin identified denominator exclusions by reviewing the CT thorax study claims for diagnoses associated with exclusion conditions.

1.3. What are the dates of the data used in testing? January 2013 – December 2013

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: state, national	☒ other: state, national

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The testing sample included 3,666 facilities that met minimum case count requirements. The number of measured entities (hospital outpatient facilities) varies by testing type; see Section **1.7** for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
The number of patients varies by testing type; see Section **1.7** for details. Prior to applying minimum case count and exclusion criteria, there were 1,669,952 CT thorax study denominator cases in the hospital outpatient (facility) setting associated with 1,270,899 patients.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.
There is no sampling; 100% Medicare fee for service claims support the measure´s calculation. The measure was originally developed and pilot tested in 2007 with a focus on Medicare eligible persons receiving CT Thorax procedures in hospital outpatient associated facilities. There are no exclusions for this measure that necessitated identifying and/or classifying patients with distinguishing diagnoses that would warrant routine appropriate use of a "combined" CT Thorax study. The focus of this measure has remained unchanged since CMS proposed its inclusion in the Hospital Outpatient Quality Reporting Program in 2009, with public reporting commencing in summer 2010.

The data sources, dates, number of measured entities, number of CT thorax studies, number of combined CT thorax studies, level of analysis, and demographic profile for the patients used in each type of testing are as follows:

Reliability Testing
Data Source: Denominator: SAF-O; Numerator: SAF-O; Exclusions: SAF-O
Dates: Denominator: January 1, 2013-December 31, 2013; Numerator: January 1, 2013-December 31, 2013; Exclusions: January 1, 2013-December 31, 2013
Number of Measured Entities: 3,666 Facilities
Number of CT Thorax Studies: 1,640,642
Number of Combined CT Thorax Studies: 39,549
Level of Analysis: Facility
Patient Characteristics: Gender (% Male): 45.9; Mean Age (Years): 71.1 (St. Dev.: 10.7); Race/Ethnicity (% Minority): 13.9

Validity Testing
Data Source: Structured qualitative survey questions completed by Technical Expert Panel (TEP) members
Dates: June 2015
Number of Responses: 10
Respondent Characteristics: CORE and Lewin asked respondents to select at least one of the following categories: insurer/purchaser (3); payer (1); clinician (5); management/administration (5); patient/patient advocate/caregiver (3).

Exclusions Analysis
Data Source: Denominator: SAF-O; Numerator: SAF-O; Exclusions: SAF-O
Dates: Denominator: January 1, 2013-December 31, 2013; Numerator: January 1, 2013-December 31, 2013; Exclusions: January 1, 2013-December 31, 2013
Number of Measured Entities: 3,664 Facilities
Number of CT Thorax Studies: 1,644,708
Number of Combined CT Thorax Studies: 39,890

Level of Analysis: Facility
Patient Characteristics: Gender (% Male): 45.9; Mean Age (Years): 71.1 (St. Dev.: 10.7); Race/Ethnicity (% Minority): 13.9

Risk Adjustment/Stratification
N/A

Identification of Statistically Significant & Meaningful Differences in Performance
Data Source: Denominator: SAF-O; Numerator: SAF-O; Exclusions: SAF-O
Dates: Denominator: January 1, 2013-December 31, 2013; Numerator: January 1, 2013-December 31, 2013; Exclusions: January 1, 2013-December 31, 2013
Number of Measured Entities: 3,666 Facilities
Number of CT Thorax Studies: 1,640,642
Number of Combined CT Thorax Studies: 39,549
Level of Analysis: Facility
Patient Characteristics: Gender (% Male): 45.9; Mean Age (Years): 71.1 (St. Dev.: 10.7); Race/Ethnicity (% Minority): 13.9

Comparability of Performance Scores when More than One Set of Specifications
N/A

Missing Data Analysis and Minimizing Bias
N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? *For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).*
We assessed patient-level SDS factors as part of the regression model reported in Section **1b.4**, which provides an overview of disparities in care for patient sub-populations. CORE and Lewin based this analysis on SDS variables included in the CMS Patient Eligibility file:

- Age group
- Gender
- Race

While an analysis of SDS factors is important in understanding differences in care for patient sub-populations, this measure is a process measure that is neither risk-adjusted nor risk-stratified. We determined that risk adjustment and risk stratification were not appropriate based on the current evidence base and the measure construct. Additional information on this determination is provided in Section **2b4.2**.

2a2. RELIABILITY TESTING
Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
Results are not based on a sample, but on 100% of fee for service Medicare claims data. In February-March 2010, CMS conducted a dry run of four outpatient imaging efficiency measures including OP-11: Thorax CT Use of Contrast.
The goals of the dry run were to (1) educate hospitals about outpatient imaging efficiency measures; (2) test the CMS measure production process; and, (3) give hospitals an opportunity to provide CMS with their feedback on

the measures. All hospitals with any outpatient hospital department claims data for the measure were included in the dry run analyses.
The measure´s dry run constituted reliability testing as it demonstrated that data elements supported by Medicare claims data were accurate across the entities being measured and helped establish measure-specific precision estimates (i.e., minimum case counts).

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

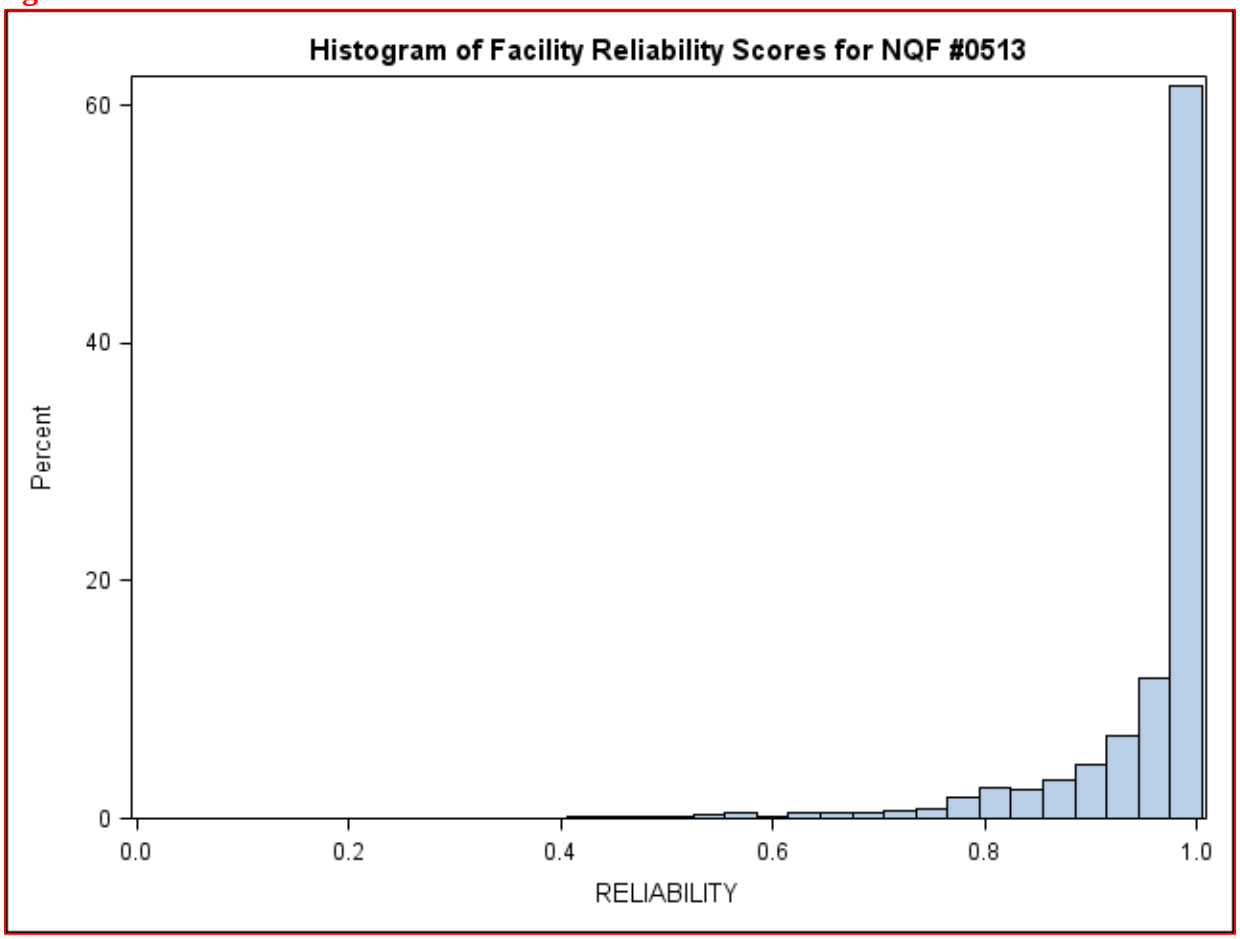
2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
To assess reliability during the dry run, CMS sent to hospitals their hospital specific report and related patient level data via the secured exchange of the QualityNet website. Hospitals were then able to compare numerator and denominator results with their internal diagnostic imaging utilization statistics, and for patients who received "combined" scans, patient records.

We calculate reliability in a manner consistent with NQF guidance and in accordance with the methods discussed in *The Reliability of Provider Profiling: A Tutorial* (2009). The reliability testing calculates the ability of the measure to distinguish between the performance of different facilities. Specifically, the testing calculated the signal-to-noise ratio for each facility meeting the minimum case count in 2013. CORE and Lewin estimate the reliability score using a beta-binomial model, which is appropriate for the reliability testing of pass/fail measures. The reliability score for each facility is a function of the facility's sample size and score on the measure, and the variance across facilities.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing?
(e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)
The dry run included 3354 hospitals that downloaded at least their hospital specific reports and 3060 hospitals that downloaded at least their patient level data, with 3007 hospitals downloading both. During the dry run process, 540 emails were submitted containing a total of 583 questions and/or comments about the four imaging efficiency measures. Regarding the CT Thorax measure, only 12 comments were received (i.e., 2% of all comments).
From these findings, and the subsequent low level of inquiries about the technical specification of this measure, CMS infers that Medicare claims data has provided reliable results about the measure´s numerator and denominator values for the over 3000 hospitals participating in Hospital Compare.

Figure 1 (below) is a histogram of the distribution of the reliability scores for the facilities meeting the minimum case count in 2013. Reliability scores ranged from 30.3% to 100.0%, with a median reliability score of 99.0%.

Figure 1



2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Calculated using a beta-binomial model, a median reliability score of 99.0% is indicative of strong measure reliability. The results of this test indicate that the measure is able to identify true differences in performance between individual facilities.

2b2. VALIDITY TESTING
2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
The measure specifications follow the ACR Appropriateness Criteria which consistently indicate a lack of appropriate use for a combined CT thorax study, and thus serve to establish strong consensus-driven construct validity for the measure.
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)
☒ **Performance measure score**
☐ **Empirical validity testing**
☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
To date, based on claims analysis, face validity is evident for the measure as indicated by: (1) two years of Hospital Compare public reporting, a systematic and transparent process; (2) ongoing surveillance of the

measure by CMS Imaging Efficiency Measures Technical Expert Panel; and public recognition (discussed in the Usability Section below) that performance scores resulting from the measure as specified can be used to distinguish good from poor quality (e.g., based on 2009 Medicare claims data, while 1700 facilities report performing CT Thorax with and without contrast less than 2% of the time, there are approximately 900 facilities reporting "combined" studies at least 7% of the time, and 365 facilities with 23% or more of their CT Thorax studies with and without contrast.

Face validity of the measure score was systematically assessed through survey of the Technical Expert Panel (TEP). Ten TEP members responded to the survey. Respondent perspectives include insurers/purchasers, clinicians, management or administration, patients/patient advocates, and caregivers. Prior to responding to questions related to measure-score and data-element face validity, we provided TEP members with detailed measure specifications.

The following statement related to measure-score face validity was posed to the TEP:

1. The measure helps assess the inappropriate use of thorax CT studies with and without contrast (combined scan).

Data-element face validity was also assessed, using the following questions and statements:

1. Thorax CT tests performed (those with contrast, those without contrast, and those with/ without contrast (combined scan)) can be accurately captured using claims data.
2. For NQF #0513, do you foresee any challenges in capturing any of these exclusions in claims data?
3. Certain conditions exclude patients from NQF #0513 as thorax CT studies with and without contrast may be appropriate for these patients. For this measure, indications for measure exclusion include any patients with diagnosis codes from the following categories.

Responses to Question #1 in the measure-score face-validity section and Question #1 in the data-element face-validity section were collected using a five-point scale: *strongly agree*, *agree*, *undecided*, *disagree*, *strongly disagree*, and *do not know*. For data-element face validity, responses to Question #2 were collected using *yes/no* response options; responses to Question #3 were collected using *keep/remove* response options.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Face validity. As noted above, the measure´s specification derives from a synthesis of ratings from a broad array of clinical conditions associated with the appropriate use of CT Thorax studies. Moreover, initial measure testing was conducted to look for consistencies in measure calculations between geographic locations (i.e., urban, rural, state) and hospital characteristics (i.e., teaching status, bed size).
Results of the face-validity assessment indicate that a diverse group of stakeholders, a majority of whom were not involved in the measure's development, support the validity of the measure. Results for each of the questions provided above follow.

Measure-Score Face Validity

1. *The measure helps assess the inappropriate use of thorax CT studies with and without contrast (combined scan).*

Response Option	Response (%)	Response (#)
Strongly Agree	40	4
Agree	50	5
Undecided	0	0
Disagree	0	0
Strongly Disagree	0	0

Do Not Know or Not Applicable	10	1

Data-Element Face Validity

1. *Thorax CT tests performed (those with contrast, those without contrast, and those with/ without contrast (combined scan)) can be accurately captured using claims data.*

Response Option	Response (%)	Response (#)
Strongly Agree	50	5
Agree	40	4
Undecided	0	0
Disagree	0	0
Strongly Disagree	0	0
Do Not Know or Not Applicable	10	1

2. *For NQF #0513, do you foresee any challenges in capturing any of these exclusions in claims data?*

Response Option	Response (%)	Response (#)
Yes	0	0
Not Sure or Do not Know	40	4
No	60	6

3. *Certain conditions exclude patients from NQF #0513 as thorax CT studies with and without contrast may be appropriate for these patients. For this measure, indications for measure exclusion include any patients with diagnosis codes from the following categories.*

Response Option	Keep this Exclusion	Remove this Exclusion	Do not Know or Not Applicable
Internal Injury of Chest, Abdomen, and Pelvis	6	0	4
Injury to Blood Vessels	6	0	4
Crushing Injury	6	0	4

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

TEP members evaluated the face validity of NQF #0513 at the data-element and measure-score level. For questions where TEP members provided a response of *Do Not Know*, respondents did not feel they had the clinical knowledge to provide a definitive response. Ninety percent of respondents believe this measure helps assess the inappropriate use of thorax CT studies with and without contrast (combined scans). Ninety percent of respondents also believe that NQF #0513 accurately captures Thorax CT tests performed (those with contrast, those without contrast, and those with/without contrast (combined studies)) through claims data. Sixty percent do not foresee any challenges capturing exclusion through claims data. The TEP members believe that none of the exclusion conditions should be removed.

Historically, measures that rely on claims data for calculation of performance are assumed to have strong face validity. This assumption of face validity is due in part to the rigor with which data are cleaned and audited prior to payment and subsequent use in measure calculation. For other public reporting programs for which payment is adjusted based on provider performance, few concerns about use of claims data for face validity have been raised.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)
Results are not based on a statistical sample but on 100% of occurrences. There are no exclusions.

CORE and Lewin tested measure exclusions to determine the prevalence of each exclusion, by facility, and at an aggregate level. We also tested the effect of all exclusions to determine the total effect of measure exclusions on performance, both by reporting summary statistics and by calculating a spearman rank correlation coefficient. The analysis tested the following categories of measure exclusions in 2013 performance data:

- Internal Injury of Chest, Abdomen, and Pelvis
- Injury to Blood Vessels
- Crushing Injury

Currently, the measure excludes patients with any one of the above-listed conditions.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
N/A

CORE and Lewin examined overall frequencies and proportions of denominator cases excluded for each exclusion, among all CT thorax studies, for a sample of 3,664 facilities meeting the minimum case count requirements in 2013, imposing no measure exclusions. The initial patient population included 1,644,708 CT thorax studies in facilities. The total number of exclusion occurrences exceeded the number of excluded cases because a single beneficiary might meet multiple exclusion criteria.

Facilities					
Exclusion	*Overall Occurrence (N)*	*Overall Occurrence (%)*	*Distribution Across Facilities (%)*		
			25th	*50th*	*75th*
Internal Injury of Chest, Abdomen, and Pelvis	3,501	0.21	0.00	0.07	0.52
Injury to Blood Vessels	60	<0.01	0.00	0.00	0.00
Crushing Injury	57	<0.01	0.00	0.00	0.00
All Current Exclusions	3,591	0.22	0.00	0.09	0.54

Additionally, we calculated descriptive statistics for the measure scores of each facility, with and without exclusions.

Facilities		
Descriptive Statistic	*With Exclusions (%)*	*No Exclusions (%)*
Minimum	0.00	0.00
Maximum	56.84	56.84
Mean	3.96	3.95

Standard Deviation	6.95	6.93
25th Percentile	0.17	0.17
50th Percentile	1.22	1.21
75th Percentile	4.39	4.36

Finally, CORE and Lewin calculated a Spearman rank correlation coefficient for facility score with and without exclusions:

$r_S = 0.9995$
$p = 0.0000$

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
The frequency of exclusions varied across facilities (IQR: 0%-0.54%), however, the prevalence of exclusions for the denominator population was very low. While median performance does not change significantly, exclusion of patients with a crushing injury or injury to the chest, abdomen, pelvis, or blood vessels is necessary to ensure the face validity of the measure and to align with clinical guidelines. As noted in the American College of Radiology (ACR) Appropriate Use Criteria for Blunt Abdominal Trauma (2012), combined CT thorax studies are appropriate for stable patients with blunt abdominal trauma.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Risk adjustment was determined not to be necessary as guidelines did not indicate further need for case mix adjustments.

This measure is a process measure for which risk adjustment or risk stratification are not necessary. We determined risk adjustment and risk stratification were not appropriate based on the current evidence base and the measure construct. During the measure development and maintenance process, we perform an annual review of the literature, which included a scan for potential patient subpopulations for which there are differences in the clinical decision to perform combined CT thorax studies, absent red flag conditions; this review identified no clear evidence of an empirical relationship between SDS and facility-level measure performance.

In addition to the literature review, stakeholder feedback obtained during implementation and public reporting has not identified concerns related to SDS factors and need for risk adjustment. This supports the conceptual model upon which the measure is based. As a process-of-care measure, SDS factors should not influence the decision to image a patient with and without contrast materials; rather, adjustment would risk masking such

important disparities in care delivery. Variation across patient populations is reflective of differences in the quality of care provided to the patient sub-population included in the measure's denominator.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)
N/A

CORE and Lewin did not perform risk adjustment or stratification.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
N/A

CORE and Lewin did not perform risk adjustment or stratification.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
CORE and Lewin did not perform risk adjustment or stratification.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9
This measure is not risk adjusted or risk stratified.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
This measure is not risk adjusted or risk stratified.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
This measure is not risk adjusted or risk stratified.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
This measure is not risk adjusted or risk stratified.

2b4.9. Results of Risk Stratification Analysis:
This measure is not risk adjusted or risk stratified.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.e., *what do the results mean and what are the norms for the test conducted*)
This measure is not risk adjusted or risk stratified.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
This measure is not risk adjusted or risk stratified.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Descriptive analysis of Medicare claims data based on hospital outpatient facilities meeting a minimum case count criterion. For a description of CMS´ minimum case count methodology for outpatient imaging efficiency measures including this measure, see https://www.cms.gov/HospitalQualityInits/Downloads/HospitalOutpatientImagingEfficiencyMinimumCaseCounts.pdf

Among measured entities that perform only a handful of CT thorax studies, one or two cases could significantly influence and/or skew the results of the measure. Therefore, CMS applies minimum case count requirements before reporting performance scores for measured entities. The minimum case count requirements applied for this measure and other imaging efficiency measures assure a 90% confidence level for the observed rate.

For OP-11, we use two different processes for determining required case counts depending on whether the performance rate is less than 5% or greater than 95% (i.e., towards the end of the range of possible rate values), or somewhere between 5% and 95% (inclusive). Each process has three steps: (1) determine reasonable levels of precision; (2) determine the level of confidence to be required for the measures; and, (3) calculate the case counts needed to meet the precision requirements. For performance rates less than 5% or greater than 95%, we calculated the case count needed to attain the required precision to be 45 cases. For performance rates between 5% and 95%, the case count needed to attain the required precision ranges from 31 to 67 cases. This composite process for setting the minimum case count requirements optimizes precision while also maximizing the number of reporting hospitals.

A more detailed presentation of the methodology explaining the minimum case count calculations for this measure and the other outpatient imaging efficiency measures can be found at https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228889854907&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3D2012_OIE_MCC.pdf&blobcol=urldata&blobtable=MungoBlobs.

Following the application of the minimum case count, we also tested the statistical significance of the difference between facility performance scores and the mean performance value. For the 2013 data, this included 3,666 facilities. For each facility, the facility performance score and standard deviation was calculated. This analysis identified 190 (5.2%) facilities as statistical outliers. Additional details of this analysis are provided in Section **2b5.2**.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance as follows:

Facilities: 3,652
25th percentile: .005
Median: .020

75th percentile: .071
90th percentile: .232
95th percentile: .400
Thus, 913 facilities performed less than 1 "combined" CT Thorax studies per 100 CT Thorax procedures, while 365 facilities performed 23 such studies per 100 CT Thorax procedures.

Of the 3,666 facilities in 2013 meeting the minimum case count, 190 (5.2%) facilities had a performance value that was statistically significantly different from the weighted mean (or benchmark value). Statistically meaningful difference was defined as when the facility score fell outside of the confidence interval (± 1.96 standard deviations) for the measure mean (benchmark value). Thus, this calculation identifies statistical outliers.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Analysis of the 2013 performance data, and the subsequent rate of identification of statistically different performance for 5.2% of measured entities, demonstrates the ability of the measure to identify outlying performance. While many facilities have converged around a performance score of between 1% and 4%, more than 5% of facilities continue to have outlying performance. By reporting a measure mean (benchmark value), this provides an opportunity for outlying facilities to identify their high rate of overuse and work to implement quality improvement mechanisms to reduce the rate of overuse of combined CT thorax studies.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
No sampling. The measure relies on 100% fee for service Medicare claims. Multiple scores of data are not used.

This measure only uses one set of specifications.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

This measure only uses one set of specifications.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e.*, *what do the results mean and what are the norms for the test conducted*)
N/A

This measure only uses one set of specifications.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
This measure is calculated from claims data submitted by measured entities for purposes of payment. The administrative claims data used to calculate the measure are maintained by CMS's Office of Information Services; these data undergo additional quality assurance checks during measure development and maintenance. Thus, the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
As described in Section **2b7.1**, the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e.*, *what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
As described in Section **2b7.1**, the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data. As such, missing data does not bias the performance results.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes
For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources
The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
No feasibility assessment **Attachment:**

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
This measure is claims based and uses CMS hospital outpatient claims as its data source.

Special attention needs to be taken when counting procedures on the Medicare claims files. The biggest issue is how to deal with modifier codes. Modifiers are two digit indicators (alpha or numeric) that represent a service or procedure that has been altered by some specific circumstance, which typically will impact the payment amount.

Procedure modifier code “26” represents the professional component of a procedure and includes the clinician work (i.e., the reading of the image by a physician), associated overhead and professional liability insurance costs. This modifier corresponds to the human involvement in a given service or procedure.

The procedure modifier code “TC” represents the technical component of a service or procedure and includes the cost of equipment and supplies to perform that service or procedure. This modifier corresponds to the equipment/facility part of a given service or procedure.

In most cases, unmodified codes represent a global procedure, which includes both the professional and technical components. There are also other modifier codes. All other modifier codes have been counted as a technical code for our purposes. When calculating the measures, we are only concerned with procedures associated with technical and global modifiers, as these modifiers refer to services provided by the facility. This reduces the possibility of double-counting procedures, since a single procedure may result in both a technical and professional record on the claims files. There were very few instances when this occurred as it is related to procedures applicable to the measure.

When developing counts of procedures, the objective is to avoid double-counting procedures that may have been billed through multiple revenue centers within a facility. Billing through multiple centers leads to multiple records in the Medicare claims files (i.e., the SAFs). For instance, there may be multiple bills for a single contrast CT study. On one bill, the charges relate to the application of a radiopharmaceutical, which could have a technical modifier code and come from the pharmacy revenue center. On the other bill, the charges relate to the imaging study and may fall under a technical bill from the imaging center revenue center. In this case, we only count the contrast CT once, since only one contrast CT study was performed. However, if we were summing up the Medicare paid amounts for this procedure, we would include the Medicare paid amounts from both bills, as they each represent payments for services directly related to the particular contrast CT procedure.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
No fees, licensure, or other requirements are necessary to use this measure; however, CPT codes, descriptions, and other data are

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Outpatient Quality Reporting http://www.medicare.gov/hospitalcompare/search.html Hospital Outpatient Quality Reporting https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120 Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospital Outpatient Quality Reporting Hospital Outpatient Quality Reporting http://www.medicare.gov/hospitalcompare/search.html https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting:
Name of program and sponsor: The CMS Hospital Outpatient Quality Reporting (HOQR) Program
Purpose: The HOQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the HOQR Program provides CMS with data to help Medicare beneficiaries make more informed decisions about their health care. Hospital quality of care information gathered through the HOQR Program is publicly available on the Hospital Compare website.
Accountable entities and patients: The publicly reported values (on Hospital Compare) are calculated for all facilities in the United States that meet minimum case count requirements. For the period of 2010 to 2015, 2,413 facilities met the minimum case count each year. Additional facilities met the minimum case count requirements in some, but not all, years. The claims included in the publicly reported calculations are for Medicare FFS patients whose claims are subject to the Outpatient Prospective Payment System (OPPS).

Quality Improvement with Benchmarking (external benchmarking to multiple organizations):

Name of program and sponsor: The CMS HOQR Program
Purpose: The HOQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the data is publicly reported on the Hospital Compare website. The data reported on Hospital Compare not only shows the hospital's score on the measure, but also provides state and national averages for the measure. This enables consumers to compare the hospital's performance to other facilities and determine if the facility is performing well or not; providing the consumer with comparative information will help them decide where to seek care.
Accountable entities and patients: The publicly reported values (on Hospital Compare) are calculated for all facilities in the United States that meet minimum case count requirements. For the period of 2010 to 2015, 2,413 facilities met the minimum case count each year. Additional facilities met the minimum case count requirements in some, but not all, years. The claims included in the publicly reported calculations are for Medicare FFS patients whose claims are subject to the OPPS.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This measure is publicly reported.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
This measure is publicly reported.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Summary statistics for performance scores from 2010 - 2015 are provided in Section 1b.2.

The median rate of overuse decreased significantly from 2010 to 2015. The median rate of overuse declined by 47.4% (1.9% to 1.0%). Over the period of 2010 to 2015, 2,413 facilities met the minimum case count to be eligible for public reporting in all years. Additional facilities met the minimum case count requirements in some, but not all, years. During the 2010 public reporting period, there were more than 76,950 inappropriate CT thorax combined studies performed for Medicare FFS beneficiaries in the United States. This number fell to approximately 39,954 potentially inappropriate CT thorax combined studies in the most recent year of publicly reported data.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable as there is demonstrated improvement in measure performance.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure testing. Similarly, no evidence of unintended consequences to individuals or populations have been reported since implementation. We will continue to monitor the potential for unintended consequences through an annual review of the literature as well as an ongoing review of stakeholder comments and inquiries.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)
Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.Meyyur@cms.hhs.gov, 410-786-7224-
Co.3 **Measure Developer if different from Measure Steward:** The Lewin Group
Co.4 **Point of Contact:** Colleen, McKiernan, Colleen.McKiernan@lewin.com, 703-269-5595-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

The contractor has convened a TEP, which will evaluate and provide feedback on measure-development and maintenance efforts for the imaging efficiency measures. Specifically, the TEP will provide direction and feedback through all phases of project activities, including expansion of imaging efficiency measures to additional CMS quality reporting programs, updates to the current specifications of the seven imaging efficiency measures, review of quantitative testing results, feedback on qualitative testing questions (i.e., results of TEP member questionnaires), and support for endorsement of the measures by the National Quality Forum (NQF).

The following is a list of the contractor's TEP members:

Meenu Arora, MBA
Quality Improvement Leader , Sequoia Hospital

Brian Baker
Chief Executive Officer, Carealytics

Peter Benner
Vice Chair, MNSure

Martha Deed, Ph.D
Safe Patient Project´s Patient Advocacy Network

Lawrence Feinberg, MD
Attending Physician, University of Colorado Hospital

Elliott Fishman, MD
Professor of Radiology and Oncology, Johns Hopkins School of Medicine

Marian Hollingsworth
Patient Advocate

Michael Hutchinson, MD, Ph.D
Clinical Associate Professor of Neurology, Icahn School of Medicine at Mount Sinai

Gregory M. Kusiak, MBA, FRBMA
President, California Medical Business Services, Inc.

Barbara Landreth, RN, MBA
Clinical Information Analyst , St. Louis Area Business Health Coalition

Barbara McNeil, MD, Ph.D
Head Professor of Radiology, Harvard University

Michael J. Pentecost, MD
Chief Medical Officer, NIA Magellan

David Seidenwurm, MD
Medical Staff Consultant, Sutter Medical Group

Adam Sharp, MD, MS
Research Scientist , Kaiser Permanente Southern California

Paul R. Sierzenski, MD, RDMS, FACEP, FAAEM

Medical Director, Christian Health Care System

C. Todd Staub, MD, FACP
Chairman, ProHealth Physicians

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2009

Ad.3 **Month and Year of most recent revision:** 03, 2015

Ad.4 **What is your frequency for review/update of this measure?** Annually

Ad.5 **When is the next scheduled review/update for this measure?** 03, 2016

Ad.6 **Copyright statement:** This measure does not have a copyright.

Ad.7 **Disclaimers:** CPT codes, descriptions, and other data only are copyright 2013 American Medical Association (AMA). All rights reserved. CPT is a registered trademark of the AMA. Applicable Federal Acquisition Regulation Site (FARS)\Defense Federal Acquisition Regulation Statement (DFARS) Restrictions Apply to Government Use. Fee schedules, relative value units, conversion factors and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0577
De.2. **Measure Title:** Use of Spirometry Testing in the Assessment and Diagnosis of COPD
Co.1.1. **Measure Steward:** National Committee for Quality Assurance
De.3. **Brief Description of Measure:** The percentage of patients 40 years of age and older with a new diagnosis of COPD or newly active COPD, who received appropriate spirometry testing to confirm the diagnosis.
1b.1. **Developer Rationale:** This measure assesses whether patients considered to have a diagnosis of COPD through the presence of symptoms and risk factors (e.g., dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease) had a spirometry assessment to confirm the diagnosis. The improvement in quality envisioned by the use of this measure is to ensure that patients receive spirometry testing to confirm a COPD diagnosis and determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to guide therapy.

S.4. **Numerator Statement:** At least one claim/encounter for spirometry during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date. The Index Episode Start Date is the earliest date of service for an eligible visit (outpatient, ED or acute inpatient) during the 6 months prior to the beginning of the measurement year through 6 months after the beginning of the measurement year with any diagnosis of COPD.
S.7. **Denominator Statement:** All patients age 42 years or older as of December 31 of the measurement year, who had a new diagnosis of COPD or newly active COPD during the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year.
S.10. **Denominator Exclusions:** N/A

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Electronic Clinical Data
S.26. **Level of Analysis:** Health Plan, Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: Dec 04, 2009 **Most Recent Endorsement Date:** Jul 31, 2012

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the measures still meet the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- During the previous review, the Committee agreed the evidence was appropriate and consistent for the use of spirometry to confirm the diagnosis of Chronic Obstructive Pulmonary Disease (COPD), however, it expressed concerns about the developer's assessment that confirming the diagnosis improves overall outcomes.
- Updated evidence for this process measure is based on three clinical practice guidelines for the diagnosis and management of Chronic Obstructive Lung Disease.
  - Dated 2015, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines provide ungraded recommendations for COPD assessment. The GOLD guidelines referenced 613 studies to update the previous set of guidelines from 2013. The recommendation for spirometry to confirm a COPD diagnosis was based on a systematic review and three observational studies.
  - The American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society 2011 guidelines graded the evidence as a strong recommendation with moderate quality evidence (the second highest ranking in this grading system). The recommendation is: "spirometry should be obtained to diagnose airflow obstruction in patients with respiratory symptoms." The recommendation for spirometry to confirm a COPD diagnosis cited 17 randomized control trials, meta-analyses, systematic reviews, and observational studies.
  - The Institute for Clinical Systems Improvement (ICSI) Guidelines, dated 2013, referenced a systematic review related to spirometry testing to confirm a COPD diagnosis. No grading.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates: Guidelines republished since last review. The 2015 GOLD guidelines referenced 613 studies to update the previous set of guidelines from 2013. The 2011 American College of Physicians et al. guidelines referenced 62 studies to update the previous set of guidelines from 2007.

Exception to evidence: Not applicable

Guidance from the Evidence Algorithm: 1→3 →4→5 (eligible for HIGH rating)

Questions for the Committee:

- *Is the timeframe of 2 years prior to the Index Episode Start Date through 6 months after the Index Episode Start Date evidence-based? If not, is it reasonable (i.e., have face validity)?*
- *Is the new evidence sufficient to demonstrate that confirming the diagnosis of COPD improves overall outcomes?*
- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The previous Committee agreed the data demonstrated underutilization of spirometry with mean results at the health plan level corresponding to 41.7% (2010), 38.8% (2009) and 37.6% (2008).
- Data were collected from the NCQA Healthcare Effectiveness Data and Information Set (HEDIS) for Commercial HMOs and PPOs, Medicare HMOs and PPOs, and Medicaid HMO. The mean results ranged from 31% to 44% amongst the various plans with little change seen from 2012 to 2014 (~1%) within each plan type.

Disparities

- NCQA does not currently collect performance data stratified by race, ethnicity, or language. While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Do the data demonstrate a potential for improvement in COPD prevention and disease management?*
- *Since no disparities information are provided, are you aware of evidence that disparities exist in this area of healthcare?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**Process outcome based on claims/electronic data for health plan/system level of analysis.
Focuses on Spirometry for COPD Diagnosis (different than 0091).
Guidance from the Evidence Algorithm: 1→3 →4→5 (eligible for HIGH rating)
Questions for the Committee:
o Is the timeframe of 2 years prior to the Index Episode Start Date through 6 months after the Index Episode Start Date evidence-based? If not, is it reasonable (i.e., have face validity)? YES
o Is the new evidence sufficient to demonstrate that confirming the diagnosis of COPD improves overall outcomes? YES--based on GOLD recommendations
o The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence? YES
**Process measure based on high quality evidence.
**Additional information suggests stronger relationship. However, the time frame (2 years prior; 6 months after) is a bit confusing - may need discussion because of this
**as best as I can tell the support for this is that it is listed in more guidelines since last considered. The word "should" in the ACXP guidelines may be misleading. In the regulatory space the word :should means it is a good idea but not required. The words "must" and "Shall" are required words
**The timeframe of 2 years prior to the Index Episode Start Date through 6 months after the Index Episode Start Date does not appear to be evidence-based. However, it appears to be reasonable.
The new evidence provided indicates that accurate diagnosis has the potential to optimize patient management and improve patient outcomes. However, evidence demonstrating that confirming the diagnosis of COPD improves overall patient outcomes did not appear to be provided.
I agree there is no need for repeat discussion and vote on Evidence.

1b. Performance Gap

Comments:

**Evidence shows under use of spirometry
Questions for the Committee:
o Is there a gap in care that warrants a national performance measure? At 41%--YES
o Do the data demonstrate a potential for improvement in COPD prevention and disease management? YES FOR DIAGNOSIS, NOT FOR PREVENTION.
o Since no disparities information are provided, are you aware of evidence that disparities exist in this area of healthcare? NO
**yes. Performance gap exists.
**Data suggests gap in performance, and that there is some variation between plans. Govt plans show lower

performance (Medicaid lowest). Almost no change in data over three years (flag re potential improvement?)
**The data indicate a gap in care that warrants a national performance measure.
The data provided demonstrate a potential for improvement in COPD prevention and disease management due to accurate diagnosis of the severity of the patient's COPD.
I am unaware of evidence that disparities exist in this area of healthcare.

1c. High Priority (previously referred to as High Impact)
Comments:
**Not applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data

Specifications:

- Specifications were not updated since the last review.
- The numerator for this measure is at least one claim/encounter for spirometry during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date". The denominator is all patients age 42 years or older as of December 31 of the measurement year, who had a new diagnosis of COPD or newly active COPD during the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year.
- The CPT, ICD-9, and ICD-10 codes are described in the numerator and denominator details, and included in the value sets excel attachment.
- The calculation algorithm is stated in S.18 and appears straightforward.

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- During the previous review, NCQA provided reliability statistics for this measure using HEDIS health plan performance data for 2010. The beta-binomial method was used to determine the ratio of signal to noise with results of 0.94 for Commercial, 0.92 for Medicaid and 0.97 for Medicare.

Describe any updates to testing

- NCQA updated measure score reliability and conducted new construct validity using data from all health plans that submitted HEDIS data for this measure in 2012 and 2015. The new measure score reliability and construct validity testing results used health plan data from July 1, 2011 through December 31, 2014. These analyses included all of the health plans (365 Commercial, 355 Medicare, and 124 Medicaid).

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing: Beta-binomial method was used. The developer states this method provides a better fit

when estimating the reliability of pass/fail rate measures, as is the case with most HEDIS® measure; it accounts for the non-normal distribution of performance within and across accountable entities. The developer also noted reliability scores vary from 0.0 to 1.0: A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Results of reliability testing: The developer states that a minimum reliability score of 0.7 is generally used to indicate sufficient signal strength to discriminate performance between accountable entities.

Beta-Binomial Statistic:

Commercial			Medicare			Medicaid		
Median	Overall	10th-90th	Median	Overall	10th-90th	Median	Overall	10th-90th
0.88	0.83	0.58-0.98	0.95	0.92	0.79-0.99	0.88	0.85	0.64-0.96

Guidance from the Reliability Algorithm: 1→2→4→5→6 (eligible for HIGH rating)

Questions for the Committee:

- *The specifications have not changed since the last review. The developer has provided testing at the score level using newer data. Does the Committee agree there is no need for repeat discussion and vote on Reliability?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Is the timeframe of 2 years prior to the Index Episode Start Date through 6 months after the Index Episode Start Date evidence-based? If not, is it reasonable (i.e., have face validity)?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- For the initial field test in 2004, NCQA tested the measure for face validity. The expert panels agreed that the performance on the spirometry measure was an accurate representation of quality performance and distinguished performance among health plans.

Describe any updates to validity testing:

- Construct validity testing, using Pearson's correlations of similar measures, was conducted.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing methods:

- 2004 Face Validity: The measure was assessed for face validity with input from 3 NCQA expert panels.

- 2004 Critical Data Element Validity Testing: Validity at the data element-level was tested by comparing the presence of administrative claims codes for a new diagnosis of COPD (required to calculate the denominator) and for a spirometry test performed (required to calculate the numerator) to documentation in the medical record, which is considered the "gold standard".
- 2014 Testing Construct Validity: Pearson correlation: coefficients were calculated to estimate the strength of the association of this spirometry measure to two other COPD measures: Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator. Results are reported within a ranges from -1 and +1.

Validity testing results:

- 2004 Face Validity:
  - The expert panels agreed that the performance on the spirometry measure was an accurate representation of quality performance and distinguished performance among health plans.
- 2004 Critical Data Element Validity Testing:
  - Across 4 plans, validation of a new COPD diagnosis in the medical record was 64%, with a range of 30% to 100%. Validation was higher in the Medicare population (74%) compared to the Commercial population (60%).
- 2014 Construct Validity Testing:
  - The results indicated that the COPD measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized. The level of correlation for Medicaid plans on the spirometry measure and the pharmacotherapy management for COPD exacerbations systemic corticosteroids indicator was moderate (the correlation coefficient was 0.3), while the other correlations were weaker (0.1 or 0.2). Correlations also were weaker for Medicare plans compared to Commercial and Medicaid plans.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- No Exclusions

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Question for the Committee:

- *Is no risk adjustment appropriate?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The results indicate there is an 8-11% gap in performance between the 25th and 75th percentile performing plans across the different product lines.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

Guidance from the Validity Algorithm: 1→2→3→6→7→8 (eligible rating of HIGH)

Note: The previous testing data met the criterion and was eligible for a rating of MEDIUM under the algorithm. With performance-score level testing, the measure is now eligible for HIGH.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Specifications consistent with evidence

Question for the Committee:

- Are the specifications consistent with the evidence? YES
- Is the timeframe of 2 years prior to the Index Episode Start Date through 6 months after the Index Episode Start Date evidence-based? If not, is it reasonable (i.e., have face validity)? YES

**Specifications consistent

**Don't see reason for time frames surrounding index episode state date; contrasting evidence that the use of spirometry leads to improved care, not just diagnosis confirmation?

**The validity specifications appear to be consistent with the evidence.

The timeframe of 2 years prior to the Index Episode Start Date through 6 months after the Index Episode Start Date does not appear to be evidence-based. However, it appears to be reasonable.

2a2. Reliability Testing

Comments:

**Validity testing from 2004 (Face), 2004 (Critical element), 2014 (Construct Validity).

Questions for the Committee:

- Do the results demonstrate sufficient validity so that conclusions about quality can be made? YES
- Do you agree that the score from this measure as specified is an indicator of quality? YES

**Validity sufficient so that conclusions about quality can be made.

**only possible concern is if there is a link between measure and improved care, not just improved diagnosis

**I believe so

**The results presented by the developer appear to demonstrate sufficient validity so that conclusions about quality can be made.

I agree the score from this measure as specified is an indicator of quality.

2b2. Validity Testing

Comments:

**2b3: No exclusions

2b4: No risk adjustment

2b5: meaningful difference--8-11% gap in performance between 25th and 75th percentile

Question for the Committee:

- Does this measure identify meaningful differences about quality? YES

2b6: Comparability not applicable

2b7: Missing data--results not biased by missing data

Guidance from the Validity Algorithm: 1→2→3→6→7→8 (eligible rating of HIGH)

Note: The previous testing data met the criterion and was eligible for a rating of MEDIUM under the algorithm. With performance-score level testing, the measure is now eligible for HIGH.

**High

**Risk Adjustment: no risk adjustment appears to be appropriate

Meaningful Differences: this measure appears to identify meaningful differences in the use of spirometry for COPD diagnosis.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:

**Guidance from the Reliability Algorithm: 1→2→4→5→6 (eligible for HIGH rating)

Questions for the Committee:

- The specifications have not changed since the last review. The developer has provided testing at the score level using newer data. Does the Committee agree there is no need for repeat discussion and vote on Reliability? YES

**HIGH

**I believe so

**I agree there is no need for repeat discussion and vote on Reliability.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.
- NCQA conducts an independent audit of all 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.

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**Electronic data used for performance measurement; NCQA also conducts an independent audit of all HEDIS collection.
Criteria met.
**Electronic clinical data.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure :
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- This measure is publically reported nationally and by geographic regions in NCQA's State of Health Care annual report.
- The measure is reported in Consumer Reports and on the NCQA website and used to calculate health plan ratings.
- This measure is used in Quality Compass.

Improvement results: From 2012-2014, the average rate did not increase across Commercial, Medicare and Medicaid plans. The mean results ranged from 31% to 44% amongst the various plans with little change seen (~1%) within each plan. There was slight improvement in Medicare PPO plans at the 90th percentile.

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms: The developer states there were no identified unintended consequences for this measure during testing or since implementation.

Feedback: No feedback provided on QPS. Measure reviewed by MAP for Physician Quality Reporting System (PQRS) and Value-Based Payment Modifier Program (VBPM) in 2013, and for Physician Compare and VBPM again in 2014. MAP

recommended the developer explore creating a composite of all COPD measures and then link that composite with the COPD resource use measure.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *The data indicate there has been little change* (~1%) within each plan type over time. *Do the data indicate the measure, although in use, yields results that are not useable?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Measure is publically reported and used in accountability programs. Reported in Consumer Reports and NCQA website.
Improvement not seen in commercial, Medicare or Medicaid plans but slight change seen in Medicare PPOs at 90%tile.
Questions for the Committee:

- Can the performance results be used to further the goal of high-quality, efficient healthcare? YES
- The data indicate there has been little change (~1%) within each plan type over time. Do the data indicate the measure, although in use, yields results that are not useable? UNCERTAIN, DEFER TO PRIMARY AND SECONDARY REVIEWERS.

**Data are usable.
May wish to consider harmonizing this measure with 0102 and 0091.
**Any information on if / how measure has been used in improvement projects? Has lack of movement due to lack of focus or not wanting to change practice?
**Quality standards for office-based testing might be a concern. small offices in more remote locations may have less utilization and thus less familiarity with obtaining a high quality exam.
**The performance results can be used to further the goal of high-quality, efficient healthcare. Since this measure is currently being used in accountability programs, yet there has been little improvement in the use of spirometry for COPD diagnosis, more emphasis may need to be placed on the importance of spirometry when evaluating a patient suspected of having COPD.
The lack of change in this measure could potentially be used to improve quality in this area.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0091 : COPD: Spirometry Evaluation

Harmonization

- The previous Pulmonary/Critical Care Committee recommended that measures 0091 and 0577 be fully harmonized in order to continue endorsement. At the time, NCQA and PCPI stated the recommendations to address misalignment in the specifications were due to the different data collection and reporting environments, but they would review the differences with their respective measure advisory expert panels to harmonize, if possible.
- NCQA response to harmonization in the current submission: NQF 0091 is a physician-level measure that uses administrative claims or medical record data. There is no impact on interpretability or added burden of data collection because the data for 0057 measure is collected from different data sources by different entities and the focus of 0057 measure is different. Measure 0091 focuses on whether patients with a COPD diagnosis, not specifically a new diagnosis, had spirometry testing performed at least once during the measurement year, while 0577 specifies that patients with a new COPD diagnosis receive spirometry testing within 6 months following diagnosis.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

NOTE: evidence that was included in the 2012 submission form may not completely match with the layout of the current evidence form.

Measure Number (*if previously endorsed*)**:** 0577

Measure Title: Use of Spirometry Testing in the Assessment and Diagnosis of COPD

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: Click here to enter a date

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that

are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Spirometry testing to confirm COPD diagnosis

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Patient presents with respiratory symptoms, such as dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease >>> Provider gives patient a spirometry test and uses the spirometry test results to confirm the presence of persistent airflow limitation and thus of COPD >>> Provider also uses the spirometry test results to determine the severity of the disease and airflow limitation, the impact on the patient's health status, and the risk of future events, in order to guide therapy >>> Patients receive appropriate therapy to reduce COPD symptoms, reduce the frequency and severity of exacerbations, and improve health status and exercise tolerance (desired outcome)

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines:

Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Bethesda (MD): Global Initiative for Chronic Obstructive Lung Disease, 2015.

http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:

Qaseem A, Wilt T, Weinberger S, Hanania N, Criner G, van der Molen T, Marciniuk D, Denberg T, Schünemann H, Wedzicha W, MacDonald R, Shekelle P. Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Annals of Internal Medicine. 2011;155(3):179-191. doi:10.7326/0003-4819-155-3-201108020-00008

http://annals.org/article.aspx?articleid=479627

Institute for Clinical Systems Improvement (ICSI):

Anderson, B., K. Conner, C. Dunn, et al. 2013. Institute for Clinical Systems Improvement. Diagnosis and Management of Chronic Obstructive Pulmonary Disease (COPD). https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_respiratory_guidelines/copd/

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines:

- Page 10: A clinical diagnosis of COPD should be considered in any patient who has dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease. Spirometry is required to make the diagnosis in this clinical context; the presence of a postbronchodilator FEV1/FVC < 0.70 confirms the presence of persistent airflow limitation and thus of COPD.
- Page 12: The goals of COPD assessment are to determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to, eventually, guide therapy. To achieve these goals, COPD assessment must consider the following aspects of the disease separately:
    - Current level of patient's symptoms
    - Severity of the spirometric abnormality
    - Exacerbation risk
    - Presence of comorbidities

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:

- Recommendation 1: Recommend that spirometry should be obtained to diagnose airflow obstruction in patients with respiratory symptoms (Grade: strong recommendation, moderate-quality evidence).

ICSI Guidelines:

- Page 8: The diagnosis of COPD should be suspected based on the patient's medical history and physical examination, but spirometry is required to make a diagnosis and confirm the presence of persistent airflow limitation *(Global Initiative for Chronic Obstructive Lung Disease, 2011 [Guideline]).* Spirometry is an established and important method of measuring lung function for the diagnosis and management of patients with COPD. It is recommended for symptomatic patients at risk of COPD, particularly smokers greater than 45 years of age, and for regular follow-up of patients with documented COPD *(Wilt, 2005 [Systematic Review])*. According to the GOLD criteria, COPD is defined as an FEV1/FVC ratio less than 70% after treatment *(Global Initiative for Chronic Obstructive Lung Disease, 2011 [Guideline])*. Large population screening is not recommended.
- Page 10: It is important to distinguish COPD from asthma, because treatment and prognosis differ. Measurement of pre- and post-bronchodilator FEV1 can assist with this differentiation. In asthma, the spirometric abnormality tends to return to normal with bronchodilators, although this distinction between COPD and asthma is not strictly rigid. If the FEV1/FVC ratio improves to > 70% after bronchial dilation, a diagnosis of COPD can be ruled out. Factors commonly used to distinguish COPD from asthma include age of onset, smoking history, triggering factors and occupational history.

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

GOLD Guidelines:

The recommendations did not include grading.

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society Guidelines:

Strong recommendation/Moderate-quality evidence: Benefits clearly outweigh risks and burden or vice versa. RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise) or exceptionally strong evidence from observational studies.

ICSI Guidelines GRADE System:

- Guideline
- Systematic Review

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

GOLD Guidelines: N/A

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society Guidelines:

- Strong recommendation/High-quality evidence: Benefits clearly outweigh risks and burden or vice versa. RCTs without important limitations or overwhelming evidence from observational studies.
- Strong recommendation/Low-quality evidence: Benefits clearly outweigh risks and burden or vice versa. Observational studies or case series.
- Weak recommendation/High-quality evidence: Benefits closely balanced with risks and burden. RCTS without important limitations or overwhelming evidence from observational studies.
- Weak recommendation/ Moderate-quality evidence: Benefits closely balanced with risks and burden. RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise) or exceptionally strong evidence from observational studies.

ICSI Guidelines GRADE System Evidence Definitions:

- Meta-analysis
- Systematic review
- Decision-analysis
- Cost-effectiveness analysis
- High Quality Evidence = Further research is very unlikely to change our confidence in the estimate of effect.
- Moderate Quality Evidence = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- Low Quality Evidence = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines:

Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Bethesda (MD): Global Initiative for Chronic Obstructive Lung Disease, 2015.

http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:

Qaseem A, Wilt T, Weinberger S, Hanania N, Criner G, van der Molen T, Marciniuk D, Denberg T, Schünemann H, Wedzicha W, MacDonald R, Shekelle P. Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Annals of Internal Medicine. 2011;155(3):179-191. doi:10.7326/0003-4819-155-3-201108020-00008

http://annals.org/article.aspx?articleid=479627

Institute for Clinical Systems Improvement (ICSI):

Anderson, B., K. Conner, C. Dunn, et al. 2013. Institute for Clinical Systems Improvement. Diagnosis and Management of Chronic Obstructive Pulmonary Disease (COPD). https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_respiratory_guidelines/copd/

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

This measure assesses whether patients with a new diagnosis of COPD or newly active COPD were given a spirometry test to confirm the diagnosis. This measure is based on guidelines and evidence that spirometry should be performed to diagnose airflow obstruction in patients with respiratory symptoms in order to make a clinical diagnosis of COPD and determine appropriate therapy.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines:

Evidence was not graded for these guideline recommendations

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:

Strong recommendation/Moderate-quality evidence: Benefits clearly outweigh risks and burden or vice versa. RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise) or exceptionally strong evidence from observational studies.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:

- Strong recommendation/High-quality evidence: Benefits clearly outweigh risks and burden or vice versa. RCTs without important limitations or overwhelming evidence from observational studies.

- Strong recommendation/Low-quality evidence: Benefits clearly outweigh risks and burden or vice versa. Observational studies or case series.
- Weak recommendation/High-quality evidence: Benefits closely balanced with risks and burden. RCTS without important limitations or overwhelming evidence from observational studies.
- Weak recommendation/ Moderate-quality evidence: Benefits closely balanced with risks and burden. RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise) or exceptionally strong evidence from observational studies.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range:

GOLD Guidelines: 1965 – 2014

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society: 1966 – 2009

ICSI Guidelines: 1968-2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The guideline developers did not provide a breakdown of the specific number of randomized control trials (RCTs) and given the number of studies included in the systematic reviews we were not able to delineate all RCTs for each recommendation. The GOLD guidelines referenced a total of 613 studies to update the previous set of guidelines from 2013. The recommendation for spirometry to confirm a COPD diagnosis was based on a systematic review and three observational studies. The American College of Physicians et al. guidelines referenced a total of 62 studies to update the previous set of guidelines from 2007. The recommendation for spirometry to confirm a COPD diagnosis cited 17 randomized control trials, meta-analyses, systematic reviews and observational studies. The ICSI Guidelines referenced a systematic review related to spirometry testing to confirm a COPD diagnosis.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Although none of the guidelines included a detailed summary of the quality of the evidence, there is strong evidence from case control and other observational studies that spirometry testing improves diagnostic accuracy of COPD compared to diagnosing COPD based on symptoms alone. The quality of evidence that spirometry testing to confirm a COPD diagnosis is associated with better treatment outcomes is moderate and based on findings from RCTs. One systematic review of three RCTs that enrolled over 2,500 subjects over 3+ years found that pharmacologic treatment effectiveness was associated with disease severity as measured by baseline spirometry; therefore, spirometry testing is useful to identify those patients who might benefit from pharmacologic treatment in order to improve outcomes. The systematic review also found moderate quality evidence from five cohort studies that baseline spirometry testing provides independent prognosis of overall respiratory symptoms and morbidity and mortality in individuals with established COPD and that spirometry results may be useful as a guide for initiation of inhaled medications and pulmonary rehabilitation among individuals having symptoms, especially frequent exacerbations.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

NCQA's measure, Use of Spirometry Testing in the Assessment and Diagnosis of COPD, is based on the research literature, guidelines, and expert feedback. Though COPD is a major cause of morbidity and mortality, studies have found that the disease is under-diagnosed, particularly in its milder forms. A number of studies have found that spirometry is a valuable tool for the diagnosis of COPD. One study found that 42% of newly diagnosed cases in study participants would not have been detected without spirometry . Spirometry is particularly useful in distinguishing COPD from asthma. Major clinical guidelines designate spirometry as the gold standard for diagnosis of COPD.

On an initial visit for COPD assessment, spirometry assessments can confirm the presence and any reversibility of airflow obstruction (low FEV1 and FEV1/FVC ratio) and distinguish COPD from asthma. There is strong evidence that spirometry testing improves diagnostic accuracy of COPD compared to diagnosing COPD based on symptoms alone; for instance a systematic review found that a third of patients with normal airflow reported respiratory symptoms and 21 percent with severe airflow obstruction did not report respiratory symptoms. Another benefit of spirometry is to identify those with symptomatic, severe airflow obstruction who might benefit from pharmacologic treatment in order to improve or lessen the number of COPD exacerbations. Compared to diagnosis and treatment based on clinical examination alone, spirometry is likely to reduce the number of individuals reporting symptoms who are inaccurately diagnosed with, and treated for, COPD because they do not have airflow obstruction of severity where treatment is beneficial. The evidence supports the initiation of inhaled bronchodilator treatment (anticholinergics, long-acting β-agonists, or corticosteroids) in patients who have respiratory symptoms and FEV1 less than 60% predicted.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The physical performance of spirometry has not been associated with adverse effects. In addition, spirometry testing is relatively inexpensive. The evidence also states that spirometry testing to confirm COPD diagnosis reduces the risk of over-diagnosing and over-treating patients for COPD. There is consensus that the benefits of spirometry testing far outweigh the potential harms.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
SPR_Evidence.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure assesses whether patients considered to have a diagnosis of COPD through the presence of symptoms and risk factors (e.g., dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease) had a spirometry assessment to confirm the diagnosis. The improvement in quality envisioned by the use of this measure is to ensure that patients receive spirometry testing to confirm a COPD diagnosis and determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to guide therapy.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year and product line (i.e. commercial, Medicaid, and Medicare).

Commercial HMO Rate

YEAR	MEAN	ST DEV	MIN	10TH	25TH	50TH	75TH	90TH	MAX	Interquartile Range
2012	44%	9%	18%	34%	37%	43%	47%	57%	77%	10%
2013	42%	9%	21%	32%	37%	41%	46%	57%	76%	9%
2014	43%	10%	23%	32%	37%	42%	47%	57%	77%	10%

Commercial PPO Rate

2012	42%	6%	25%	35%	37%	41%	45%	49%	61%	8%
2013	41%	7%	23%	34%	37%	40%	45%	48%	68%	8%
2014	41%	6%	19%	35%	37%	40%	45%	50%	62%	8%

Medicare HMO Rate

2012	37%	12%	5%	23%	30%	36%	42%	51%	79%	12%
2013	36%	12%	1%	23%	28%	35%	41%	52%	81%	13%
2014	36%	13%	1%	22%	30%	35%	41%	51%	81%	11%

Medicare PPO Rate

2012	35%	9%	6%	24%	30%	35%	39%	44%	61%	9%
2013	36%	9%	20%	25%	30%	35%	40%	46%	61%	10%
2014	35%	9%	6%	26%	30%	35%	40%	46%	56%	10%

Medicaid HMO
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 31% | 9% | 8% | 19% | 26% | 31% | 38% | 43% | 66% | 12%
2013 | 31% | 9% | 10% | 20% | 26% | 30% | 37% | 42% | 60% | 11%
2014 | 31% | 8% | 11% | 21% | 26% | 31% | 36% | 41% | 55% | 10%

The data references are extracted from HEDIS data collection reflecting the most recent years of use for this measure. In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Commercial HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 186 | 495
2013 | 183 | 531
2014 | 178 | 467

Commercial PPO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 188 | 1095
2013 | 184 | 1116
2014 | 177 | 993

Medicare HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 274 | 674
2013 | 273 | 748
2014 | 257 | 763

Medicare PPO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 109 | 397
2013 | 117 | 594
2014 | 98 | 829

Medicaid HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 117 | 343
2013 | 119 | 371
2014 | 124 | 400

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
HEDIS data are stratified by type of insurance (e.g. Commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to

promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.
Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts. Health Affairs 20(10): 1984-1991.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
One study found that only a third of patients with COPD had their diagnosis confirmed with spirometry (Arne et al. 2010). Females, patients who currently smoke or have a higher BMI increased their risk of being diagnosed with COPD without fulfilling the spirometric criteria for the disease (Arne et al. 2010).

Arne M., Lisspers K., Ställberg B., Boman G., Hedenström H., Janson C., Emtner M. How often is diagnosis of COPD confirmed with spirometry? Respiratory Medicine. 2010 Apr; 104(4):550-6. doi: 10.1016/j.rmed.2009.10.023.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
The National Heart, Lung, and Blood Institute (NHLBI) states that over 15 million adults have been diagnosed with chronic obstructive pulmonary disease (COPD), and that the actual number of those with the disease may be higher (NHLBI 2013). While other major causes of death have been decreasing, COPD mortality has risen, making it the third leading cause of death in the U.S. (Hoyert and Xu 2012; NHLBI 2013). Without intervention, deaths from COPD are projected to increase by more than 30 percent in the next ten years, and COPD is projected to be the third leading cause of death worldwide (WHO 2014). In 2010, COPD was the main cause for 10.3 million physician visits, 1.5 million emergency department visits and 699,000 discharges (Ford et al, 2013). Approximately 641,000 hospital inpatient discharges and 294,000 emergency department visits in 2010 were due to COPD (CDC 2014), and the projected total cost of COPD was $49.9 billion, including $29.5 billion for direct health care expenditures (NHLBI 2009).

A spirometry test is required to confirm a COPD diagnosis and determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to guide therapy (GOLD 2015); however, sspirometry tests to confirm COPD diagnoses are largely underutilized. In a survey study of a health maintenance organization (HMO) and a university-affiliated county medical center, only 38 percent of COPD patients in the HMO and 42 percent of patients in the medical center system had spirometry results documented in their medical records (Mapel et al. 2000). A more recent study again found that only a third of patients with COPD had their diagnosis confirmed with spirometry (Arne et al. 2010).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Arne M., Lisspers K., Ställberg B., Boman G., Hedenström H., Janson C., Emtner M. How often is diagnosis of COPD confirmed with spirometry? Respiratory Medicine. 2010 Apr; 104(4):550-6. doi: 10.1016/j.rmed.2009.10.023.

Centers for Disease Control and Prevention. 2014. FastStats: Chronic Obstructive Pulmonary Disease (COPD) Includes: Chronic Bronchitis and Emphysema. http://www.cdc.gov/nchs/fastats/copd.htm (Accessed July 23, 2014).

Ford E, Croft J, Mannino D, Wheaton A, Zhang X, and Giles W. COPD Surveillance—United States, 1999-2011. Chest 2013; 144(1):284–305.

Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2015. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html (Accessed November 16, 2015).

Hoyert, D., and J. Xu. 2012. Deaths: Preliminary Data for 2011. National Vital Statistics Reports. 61(6):1-52. http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_06.pdf (Accessed July 23, 2014)

Mapel et al. Utilization in COPD: Patient Characteristics and Diagnostic Evaluation. Chest 2000; 117: 346S-353S.

National Heart, Lung, and Blood Institute. 2009. "Morbidity and Mortality: 2009 Chart Book on Cardiovascular, Lung, and Blood Diseases. https://www.nhlbi.nih.gov/resources/docs/2009_ChartBook.pdf (Accessed July 23, 2014).

National Heart, Lung, and Blood Institute. 2013. Morbidity & Mortality: 2013 Chart Book on Cardiovascular, Lung and Blood Diseases. Available at: http://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook.pdf. (Accessed November 18, 2015).

World Health Organization (WHO). 2014. Chronic Respiratory Diseases: Chronic Obstructive Pulmonary Disease. http://www.who.int/respiratory/copd/en/ (Accessed July 23, 2014).

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
N/A

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 0577_SPR_Value_Sets.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no significant changes to the measure specification since the last endorsement maintenance completed on January 31,

2012.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
At least one claim/encounter for spirometry during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date. The Index Episode Start Date is the earliest date of service for an eligible visit (outpatient, ED or acute inpatient) during the 6 months prior to the beginning of the measurement year through 6 months after the beginning of the measurement year with any diagnosis of COPD.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: A two and a half year period that begins 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date.

Denominator: A 12 month period that begins 6 months prior to the beginning of the measurement year through the 6 months after the beginning of the measurement year.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Follow the steps below to identify numerator compliance.

Identify the number of patients who had at least one claim/encounter for spirometry (Spirometry Value Set) during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date. The Index Episode Start Date is the earliest date of service for an eligible visit (outpatient, ED or acute inpatient) during the 6 months prior to the beginning of the measurement year through 6 months after the beginning of the measurement year with any diagnosis of COPD.
- For an outpatient claim/encounter, the Index Episode Start Date is the date of service.
- For an acute inpatient claim/encounter, the Index Episode Start Date is the date of discharge.
- For a transfer or readmission, the Index Episode Start Date is the discharge date of the original admission.

See corresponding Excel file for value sets referenced above.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients age 42 years or older as of December 31 of the measurement year, who had a new diagnosis of COPD or newly active COPD during the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The eligible population for the denominator is defined by following the series of steps below:

Step 1: Determine the Index Episode Start Date. Identify all patients who had any of the following during the intake period (the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year):
1) An outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set), or an ED visit (ED Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include ED visits that result in an inpatient admission.
2) An acute inpatient discharge with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). To identify acute inpatient discharges:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)

b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)
c. Identify the discharge date for the stay.

If the patient had more than one eligible visit, include only the first visit.

Step 2: Test for negative diagnosis history. Exclude patients who had any of the following during the 731-day period prior to the Index Episode Start Date.
1) An outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set), or an ED visit (ED Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include ED visits that result in an inpatient admission.
2) An acute inpatient discharge with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). To identify acute inpatient discharges:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)
c. Identify the discharge date for the stay.

For an acute inpatient Index Episode Start Date, use the Index Episode Start Date of admission to determine the 731-day period.

See corresponding Excel file for value sets referenced above.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
N/A

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
N/A

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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 measure calculation is detailed in the steps listed below:

Step 1: Determine the eligible population.

A. Determine the Index Episode Start Date. Identify all patients who had any of the following during the intake period (the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year):
1) An outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set), or an ED visit (ED Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include ED visits that result in an inpatient admission.
2) An acute inpatient discharge with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). To identify acute inpatient discharges:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)
c. Identify the discharge date for the stay.

If the patient had more than one eligible visit, include only the first visit.

B. Test for negative diagnosis history. Exclude patients who had any of the following during the 731-day period prior to the Index Episode Start Date.
1) An outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set), or an ED visit (ED Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include ED visits that result in an inpatient admission.
2) An acute inpatient discharge with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). To identify acute inpatient discharges:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)
c. Identify the discharge date for the stay.

For an acute inpatient Index Episode Start Date, use the Index Episode Start Date of admission to determine the 731-day period.

Step 2: determine the numerator. Identify the number of patients who had at least one claim/encounter for spirometry (Spirometry Value Set) during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date. The Index Episode Start Date is the earliest date of service for an eligible visit (outpatient, ED or acute inpatient) during the 6 months prior to the beginning of the measurement year through 6 months after the beginning of the measurement year with any diagnosis of COPD.
- For an outpatient claim/encounter, the Index Episode Start Date is the date of service.
- For an acute inpatient claim/encounter, the Index Episode Start Date is the date of discharge.
- For a transfer or readmission, the Index Episode Start Date is the discharge date of the original admission.

Step 3: calculate the rate: Numerator/Denominator.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
SPR_Testing.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

NOTE: testing information that was included in the 2012 submission form may not completely match with the layout of the current testing form.

Measure Number (*if previously endorsed*)**: 0577**
Measure Title: **Use of Spirometry Testing in the Assessment and Diagnosis of COPD**
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.

- **For <u>all</u> measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

<u>Note</u>: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record

☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
N/A

1.3. What are the dates of the data used in testing? Click here to enter date range

Initial testing: During measure development, we conducted a comprehensive field test in 2004 to assess feasibility of data collection and validity of performance data and critical data elements. This field test used data from measurement year 2003, which included health plan data spanning January 1, 2001 through December 31, 2003.

Systematic evaluation of face validity: The measure was tested for face validity throughout measure development from 2004 to 2006. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2006.

Measure score reliability testing: We assessed measure score reliability using data from all health plans that submitted HEDIS data to NCQA for this measure in 2011, which used data for measurement year 2010. Measurement year 2010 required health plan data from July 1, 2008 through December 31, 2010.

2015 Update: We assessed measure score reliability again and also construct validity using data from all health plans that submitted HEDIS data to NCQA for this measure in 2015, which used data for measurement year 2014. Measurement year 2014 required health plan data from July 1, 2011 through December 31, 2014.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Initial testing: To assess feasibility of data collection and validity of performance data and critical data elements, 5 Commercial health plans, 1 Medicaid health plan and 3 Medicare health plans provided individual member-level data to

NCQA for analysis. These plans were selected because they had the resources to generate the files, had sufficient sample of members with persistent asthma for analysis, and willingness to provide the data. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity: Throughout the entire measure development process from 2004-2006, the measure was tested for face validity using panels of experts with specific clinical, methodologic and operational expertise. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2006. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliation of expert panel:

1) NCQA's Respiratory Measurement Advisory Panel (RMAP) is comprised of 10 experts (8 physicians, 1 pharmacist and 1 researcher) in clinical pulmonary care, including health care providers and policy makers.

2) NCQA's Technical Measurement Advisory Panel is a 12-member panel representing health plans methodologists, clinicians and HEDIS auditors.

3) NCQA's Committee on Performance Measurement (CPM) oversees the HEDIS measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 17 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

4) [2015 Update] NCQA's HEDIS Expert Coding Panel reviewed and provided feedback on the vocabularies and definitions found in the values sets used to identify each measure component as well as the more recent mapping of ICD-9 codes to ICD-10 codes.

In 2005, the draft measure was posted for public comment, a 30-day period of review that allowed interested parties to offer feedback to NCQA about the measure. Stakeholders from various types of organizations submitted 79 comments on the measure.

Measure score reliability: Measure score reliability was calculated from Commercial, Medicaid and Medicare plans that submitted data on this measure to HEDIS in 2011.

2015 Update: Measure score reliability was calculated again and construct validity was assessed from the 365 Commercial health plans (comprising 178 HMOs and 187 PPOs), 355 Medicare plans (comprising 257 HMOs and 98 PPOs) and 124 Medicaid health plans that submitted data on this measure to HEDIS in 2015. The plans were geographically diverse and varied in size.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
Patient sample for initial measure field testing: We collected data from 5 Commercial health plans, 3 Medicare health plans and 1 Medicaid health plan to assess feasibility of data collection and validity of performance data and critical data elements. Below is a description of the sample. It includes the number of health plans that provided data for the measurement year 2003 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial	5	157
Medicare	3	414
Medicaid	1	347

2011 measure score reliability: Measure score reliability was calculated from Commercial, Medicaid and Medicare plans that submitted data on this measure to HEDIS in 2011.

2015 Update: assessed measure score reliability and construct validity: In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Data are summarized at the health plan level and stratified by product

line (i.e. commercial, Medicare, Medicaid). Below is a description of the number of health plans that submitted data for this measure to HEDIS for measurement year 2014 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial HMO	178	232
Commercial PPO	187	405
Medicare HMO	257	358
Medicare PPO	98	223
Medicaid	124	254

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.
During measure development, we conducted a comprehensive field test in 2004 to assess feasibility of data collection and validity of performance data and critical data elements using data submitted by 5 Commercial health plans, 1 Medicaid health plan and 3 Medicare health plans. The field test used the measurement year 2003 that included data from January 1, 2001 through December 31, 2003.

Face validity was demonstrated through a systematic assessment of face validity during measure development and at regular intervals since then. Per NQF instructions we have described the composition of the technical expert panel, which assessed face validity in the data sample questions above.

We assessed measure score reliability (tested using a beta-binomial calculation). These analyses included all of the health plans that submitted data for this measure to HEDIS for measurement year 2010 (July 1, 2008 through December 31, 2010).

2015 Update: the measure underwent additional analyses to assess measure score reliability (tested using a beta-binomial calculation) and construct validity (tested using Pearson's correlations of similar measures). These analyses included all of the health plans (365 Commercial, 355 Medicare, and 124 Medicaid) that submitted data for this measure to HEDIS for measurement year 2014 (July 1, 2011 through December 31, 2014).

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Measure performance results are stratified by Commercial, Medicare and Medicaid health plans.

2a2. RELIABILITY TESTING
***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

2012 Submission Form [Testing Data]: Reliability was estimated by using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan

measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped. Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

2015 Update:
In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2012 Submission Form [Testing Data]: Reliability statistics for this measure were calculated using HEDIS health plan performance data for 2010. The results are as follows:
Commercial: 0.94499
Medicaid: 0.91681
Medicare: 0.97039

2015 Update:
Beta-Binomial Statistic:

Commercial			Medicare			Medicaid		
Median	Overall	10th-90th	Median	Overall	10th-90th	Median	Overall	10th-90th
0.88	0.83	0.58-0.98	0.95	0.92	0.79-0.99	0.88	0.85	0.64-0.96

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
2015 Update:
Interpretation of measure score reliability testing:
Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The results of this beta binomial testing suggest that this measure has good reliability. The 10-90th percentile distribution of health plan level reliability on this measure show the vast majority of health plans exceeded the minimally accepted threshold of 0.7. Strong reliability is demonstrated since the majority of variances is due to signal and not to noise.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performan ce measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

METHOD OF ASSESSING FACE VALIDITY

2012 Submission Form [Testing Data]: NCQA tested the measure for face validity using a panel of stakeholders with relevant clinical expertise and research and measurement, experience. This panel included representatives from key stakeholder groups, including the CDC, pulmonologists, provider and delivery organizations and researchers (See list of current members for the Respiratory Advisory Panel (RMAP) under section Ad.1). RMAP experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2015 Update: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

This measure was developed in 2004 to assess whether patients had spirometry testing to confirm a COPD diagnosis and determine the severity of the disease and its impact on the patient's health status. NCQA and the Respiratory Measurement Advisory Panel worked together to develop the most appropriate measure for assessing spirometry testing to confirm new COPD diagnoses.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

The spirometry measure was written and field-tested in 2004. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in 2005.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

The spirometry measure was released for Public Comment in 2005 prior to publication in HEDIS. We received and responded to 79 comments on this measure. The CPM recommended moving this measure to first year data collection by a majority vote.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was

already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

The spirometry measure was introduced to HEDIS in 2005. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure to public reporting with a majority vote in 2006.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publically reported and may be used for scoring in accreditation.

The spirometry measure has been publicly reported in HEDIS since 2006.

STEP 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

The clinical guideline recommendations for spirometry testing in the assessment and diagnosis of COPD have not changed since the measure was developed in 2005; therefore, we have not made any significant changes to the spirometry measure since it was last endorsed on January 31, 2012.

Expert Participation

This measure was tested for face validity with input from three expert panels. Guidelines from the Global Initiative for Chronic Obstructive Lung Disease (GOLD) were also a strong authoritative source in applying the evidence for the spirometry measure.

We list an overview of each panel here. Please refer to Ad.1 in the submission form for the names and affiliation of experts in each panel.

1. Respiratory Measurement Advisory Panel includes 10 members (eight physicians, one pharmacist and a researcher) with expertise in respiratory care and quality measurement.
2. The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.
3. NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

ICD-10 CONVERSION:

In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA first identified value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM)to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.

3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented to for expert review and feedback.
4. NCQA RMAP clinical review: Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.
5. New value sets containing ICD-10 code recommendations were for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under **Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development**.

METHOD OF TESTING CRITICAL DATA ELEMENT VALIDITY
2012 Submission Form [Testing Data]: Validity was tested by comparing the presence of administrative claims codes for a new diagnosis of COPD (required to calculate the denominator) and for a spirometry test performed (required to calculate the numerator) to documentation in the medical record, which is considered to be the "gold standard".

METHOD OF TESTING CONSTRUCT VALIDITY
2015 Update: We tested for construct validity by exploring whether this measure was correlated with other similar measures of respiratory care. We hypothesized that organizations that perform well on the measure should perform well on other similar HEDIS measures. To test these correlations we used a Person correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

For this measure, we specifically hypothesized:
1) Performance on the spirometry measure will be positively correlated with performance on the pharmacotherapy management of COPD exacerbations bronchodilator indicator (percent of patients who were dispensed a prescription for a bronchodilator within 30 days after an acute COPD exacerbation)
2) Performance on the spirometry measure will be positively correlated with performance on the pharmacotherapy management of COPD exacerbations corticosteroid indicator (percent of patients who were dispensed a prescription for a systemic corticosteroid within 14 days after an acute COPD exacerbation)

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

RESULTS OF FACE VALIDITY ASSESSMENT
2012 Submission Form [Testing Data]: For the initial field test in 2004, we calculated the measure performance rate and discussed the validity of the performance results with the expert panels (Respiratory Measurement Advisory Panel and the Committee on Performance Measurement). Performance rates were between 30 to 32 percent among the different product lines. Performance rates were lower for men compared to women and for people ages 75+ compared to ages 40-74. The expert panels agreed that the performance on the spirometry measure was an accurate representation of quality performance and distinguished performance among health plans.

2004 Field Test: Performance Rates on the Spirometry Measure by Product Line, Age and Gender *

	Denominator	Numerator	Performance Rate
Product Line			
Commercial	3559	1141	32.1%
Medicare	1403	427	30.4%
Medicaid	347	109	31.4%

	Denominator	Numerator	Performance Rate
Age			
40-54	1759	580	33.0%
55-64	1722	563	32.7%
65-74	1029	331	32.2%
75-84	625	177	28.3%
85+	174	26	14.9%
Gender			
F	2826	947	33.5%
M	2483	730	29.4%
Total	5309	1677	31.6%

**Includes data submitted by 5 Commercial plans, 3 Medicare plans and 1 Medicaid plan using measurement year 2003*

RESULTS OF CRITICAL DATA ELEMENT VALIDTY

2012 Submission Form [Testing Data]: Across four plans, validation of a new COPD diagnosis in the medical record was 64%, with a range of 30% to 100%. Validation was higher in the Medicare population (74%) compared to the Commercial population (60%).

2004 Field Test: COPD Diagnosis Medical Record Validation by Plan and Product Line*

	Number of patients with a new COPD Dx using administrative data	% of patients that had documentation of a new COPD Dx in medical record	% of patients that did not have documentation of a new COPD Dx in medical record
Plan:			
A	72	30.0%	48.6%
B	19	68.8%	26.3%
C	20	30.0%	70.0%
D	66	100.0%	0.0%
Total	177	63.8%	31.1%
Product Line:			
Commercial	126	60.2%	32.5%
Medicare	51	73.5%	25.5%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

In four plans, there was 76.6% data consistency for spirometry testing between administrative and medical record data. This was calculated by adding the percent of spirometry tests found in administrative data and medical record data plus the percent of spirometry tests found in neither data source. Contrary to the perceived notion that many spirometry tests happen in the physician office without a claim generated, in two plans no spirometry tests were found in medical records only without a corresponding administrative claim and in the remaining two plans relatively few spirometry tests were found only in medical record data, indicating administrative data are reliable for capturing spirometry tests.

2004 Field Test: Spirometry Numerator Validation by Plan*

Plan Code	# of patients with a new COPD Dx confirmed in both admin & medical record data	% of patients with spirometry confirmed in both medical record & admin data	% of patients with spirometry confirmed in neither admin or medical record data	% of patients with spirometry confirmed in admin data only	% of patients with spirometry confirmed in medical record data only
A	15	46.7%	20.0%	33.3%	0.0%
B	11	54.5%	36.4%	9.1%	0.0%
C	5	50.0%	33.3%	0.0%	20.0%
D	66	18.2%	57.6%	4.5%	19.7%
Total	98	28.6%	48.0%	9.2%	14.3%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

RESULTS OF CONSTRUCT VALIDITY

2015 Update: The results indicated that the COPD measures were significantly (p<.05) correlated with each other in the direction that was hypothesized. The level of correlation for Medicaid plans on the spirometry measure and the pharmacotherapy management for COPD exacerbations systemic corticosteroids indicator was moderate (the correlation coefficient was 0.3), while the other correlations were weaker. Correlations were also weaker for Medicare plans compared to Commercial and Medicaid plans.

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Commercial Plans)*

	Pearson Correlation Coefficients	
	Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator
Spirometry Measure	0.2	0.2

**Includes data submitted by 241 Commercial plans to HEDIS for these measures for measurement year 2014*
Note: All correlations are significant at p<.05

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Medicare Plans)*

	Pearson Correlation Coefficients	
	Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator
Spirometry Measure	0.1	0.1

**Includes data submitted by 355 Medicare plans to HEDIS for these measures for measurement year 2014*
Note: All correlations are significant at p<.05

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Medicaid Plans)*

	Pearson Correlation Coefficients	
	Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator
Spirometry Measure	0.2	0.3

**Includes data submitted by 124 Medicaid plans to HEDIS for these measures for measurement year 2014*
Note: All correlations are significant at p<.05

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

SYSTEMATIC ASSESSMENT OF FACE VALIDITY

2015 Update: The use of spirometry in the assessment and diagnosis of COPD measure was deemed to have the desirable attributes of a HEDIS measure in 2005 (relevance, scientific soundness, and feasibility). The technical expert panels showed good agreement that the measure as specified accurately differentiates quality across providers. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2005.Our interpretation is that this measure has sufficient face validity.

CRITICAL DATA ELEMENT VALIDTY

2015 Update: The results of the critical data element validity testing demonstrate that the administrative data elements used to calculate the measure denominator (patients with a new diagnosis of COPD) and numerator (patients that had a spirometry test performed) had moderate to strong agreement with medical record data and are valid.

CONSTRUCT VALIDITY

2015 Update: Coefficients with absolute value of less than 0.2 are generally considered indicative of weak associations whereas absolute values of 0.2 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that the COPD measures are correlated with each other, suggesting they represent the same underlying quality construct of COPD quality of care.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

N/A

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

N/A

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

N/A

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2012 Submission Form [Testing Data]: We calculated performance rates for the measure as reported to NCQA as part of HEDIS health plan reporting from 2008-2010.

2015 Update: To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measured entities. However the method can be used for comparison of any two measured entities.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? *(e.g., number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined)*

2012 Submission Form [Testing Data]: The following tables detail the rates for the measure as reported to NCQA as part of HEDIS health plan reporting.

Commercial Results for numerator – Spirometry testing

	N	Mean	STDEV	STDERR	MIN	MAX	10th	25th	50th	75th	90th
2010	209	41.7	8.27	0.57	24.7	68.1	31.1	36.1	41.1	46.4	52.2
2009	224	38.8	8.82	0.59	20	83.2	28.3	33.5	38.2	42.9	50
2008	234	37.6	9.11	0.6	17.1	84.5	27.7	32	36.8	41.2	47.6

Medicaid Results for numerator – Spirometry testing

	N	Mean	STDEV	STDERR	MIN	MAX	10th	25th	50th	75th	90th
2010	95	31.3	9.87	1.01	6.79	55.9	19.1	24.6	30.5	35.5	47.2
2009	92	28.6	9.49	0.99	2.94	52.4	17.4	23.1	28	35.1	39.9
2008	78	29.3	9.66	1.09	6.81	57.1	17	22.8	28.5	34.4	42.6

Medicare Results for numerator – Spirometry testing

	N	Mean	STDEV	STDERR	MIN	MAX	10th	25th	50th	75th	90th
2010	244	33.9	10.7	0.69	5.88	64.5	20.5	26.7	33.3	40.5	46.9
2009	223	28.5	9.86	0.66	6.6	83.3	16.7	22.3	28	34.4	40.8
2008	197	27.7	9.47	0.67	5.77	81.6	16.7	22	27	32.9	38.8

2015 Update: HEDIS 2014 Variation in Performance across Health Plans

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Comm. HMO	467	43	10	32	37	42	47	57	10	<0.001
Comm. PPO	993	41	6	35	37	40	45	50	8	<0.001
Medicare HMO	763	36	13	22	30	35	41	51	11	<0.001
Medicare PPO	829	35	9	26	30	35	40	46	10	<0.001
Medicaid	400	31	8	21	26	31	36	41	10	<0.001

EP: Eligible Population, the average denominator size across all plans submitting 2014 HEDIS data for this measure
IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
2015 Update: The results above indicate there is an 8-11% gap in performance between the 25th and 75th percentile performing plans across the different product lines. For all product lines, the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance is in Medicare HMO plans, which shows an 11 percentage point gap between 25th and 75th percentile plans.

To put these meaningful differences in performance into context, we estimated that on average 84 additional members per Medicare HMO plan would have had a spirometry test performed to confirm COPD diagnosis if plans in the 25th percentile performed as well as plans in the 75th percentile. This estimate is based on the average health plan eligible population.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

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 all 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 followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will 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

In addition to the HEDIS Audit, NCQA provides a system to allow “real-time” feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular

re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Rating http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a

total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

From 2012-2014, the average rate did not increase across Commercial, Medicare and Medicaid plans (see section 1b.2 for summary of data from health plans). However, there was slight improvement in Medicare PPO plans at the 90th percentile. These data are nationally representative.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
More Medicaid plans and fewer Medicare plans reported the measure in 2014 compared to 2013 and 2012, which may help explain why the average performance rates did not substantially improve. There is hope that with increasing attention to this measure in public reporting programs, performance rates will improve.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no identified unintended consequences for this measure during testing or since implementation.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0091 : COPD: Spirometry Evaluation
0102 : COPD: inhaled bronchodilator therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
Pharmacotherapy Management of COPD Exacerbation (National Committee for Quality Assurance)

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NQF 0102 focuses on medication management for stable COPD or following an exacerbation, while our measure focuses on appropriate spirometry testing to confirm a new COPD diagnosis. There is no impact on interpretability or added burden of data collection because the focus of our measure is different. NQF 0091 is a physician-level measure that uses administrative claims or medical record data. There is no impact on interpretability or added burden of data collection because the data for our measure is collected from different data sources by different entities and the focus of our measure is different (0091 focuses on whether patients with a COPD diagnosis, not specifically a new diagnosis, had spirometry testing performed at least once during the measurement year, while 0577 specifies that patients with a new COPD diagnosis receive spirometry testing within 6 months following diagnosis).

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**

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

Respiratory Measurement Advisory Panel (RMAP) Members:
David Au, MD, MS, (CHAIR) Associate Prof. of Medicine/Investigator HSRD, Department of Veterans Affairs
Kurt Elward, MD, MPH, Clinical Professor of Family Medicine, Virginia Commonwealth University
Laura Feemster, MD, MS, Assistant Professor of Medicine, University of Washington Medical Center, VA HSR&D
Anne Fuhlbrigge, MD, Clinical Director, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School
Min Joo, MD, MPH, FCCP, Assistant Professor of Medicine, University of Illinois at Chicago/ Jesse Brown VA Medical Center
Christine Joseph, PhD, MPH, BSc, Associate Director of Research, Epidemiologist, Henry Ford Health System
Jerry Krishnan, MD, PhD, Prof. of Medicine & Public Health, Director of Population Health Sciences, AVP, Office of the VP for Health Affairs, University of Illinois Hospital & Health Sciences System
Todd Lee, PharmD, PhD, Primary: Senior Investigator, Secondary: Associate Professor, University of Illinois at Chicago
Allan Luskin, MD, Physician Pulmonologist, Healthy Airways
Richard O´Connor, MD, Director, Dept. of Quality Management, Allergist/Immunologist, Sharp Rees-Stealy Medical Group

Committee on Performance Measurement (CPM) Members:
Peter Bach, MD, Director, Health Policy Center, Memorial Sloan Kettering Cancer Center
Bruce Bagley, MD, FAAFP, Senior Advisor to the Professional Satisfaction and Practice Sustainability effort at the American Medical Association
Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna
Patrick Conway, MD, MSC, Chief Medical Officer and Deputy Administrator, Centers for Medicare and Medicaid Services
Jonathan D. Darer, MD, MPH, Chief Innovation Officer, Geisinger Health System
Helen Darling, Strategic Advisor, National Business Group of Health
Rebekah Gee, MD, MPH, FACOG, Assistant Professor, LSUHSC
Foster Gesten, MD, FACP, New York State Department of Health
Marge Ginsburg, Executive Director, Center for Healthcare Decisions
David Grossman, MD, MPH, Medical Director, Population Health
Christine S. Hunter, MD (Co- Chair), Chief Medical Officer, US Office of Personnel Management
Jeffery Kelman, MMSc, MD, Chief Medical Officer, United Stated Department of Health and Human Services
Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group
J. Brent Pawlecki, MD, MMM, Chief Health Officer, The Goodyear Tire & Rubber Company
Susan Reinhard, PhD, RN, Senior Vice President, AARP Public Policy Institute
Eric C. Schneider, MD, MSc, FACP (Co-chair), Senior Vice President, Policy and Research, The Commonwealth Fund
Marcus Thygeson, MD, MPH, Chief Health Officer, Blue Shield of California

HEDIS EXPERT CODING PANEL Members:
Glen Braden, MBA, CHCA, Attest Health Care Advisors, LLC
Denene Harper, RHIA, American Hospital Association
DeHandro Hayden, BS, American Medical Association
Patience Hoag, RHIT, CPHQ, CHCA, CCS, CCS-P, Health Services Advisory Group
Nelly Leon-Chisen, RHIA, American Hospital Association
Tammy Marshall, LVN, Aetna
Alec McLure, RHIA, CCS-P, Verisk Health
Michele Mouradian, RN, BSN, McKesson Health Solutions
Craig Thacker, RN, CIGNA HealthCare
Mary Jane F. Toomey, RN CPC, Aetna Better Health

Technical Measurement Advisory Panel Members:
Andy Amster, MSPH, Kaiser Permanente
Kathryn Coltin, MPH, Independent Consultant
Lekisha Daniel-Robinson, Centers for Medicare and Medicaid Services
Marissa Finn, MBA, Cigna HealthCare
Scott Fox, MS, MEd, Independence Blue Cross

Carlos Hernandez, CenCalHealth
Kelly Isom, MA, RN, Aetna
Harmon Jordan, ScD, RTI International
Ernest Moy, MD, MPH, Agency for Healthcare Research and Quality
Patrick Roohan, New York State Department of Health
Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC
Natan Szapiro, Independence Blue Cross

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2005

Ad.3 **Month and Year of most recent revision:** 07, 2015

Ad.4 **What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.
THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0702
De.2. **Measure Title:** Intensive Care Unit (ICU) Length-of-Stay (LOS)
Co.1.1. **Measure Steward:** Philip R. Lee Institute for Health Policy Studies
De.3. **Brief Description of Measure:** For all eligible patients =18 years old admitted to the intensive care unit (ICU), total duration of time spent in the ICU until time of discharge from the ICU; both observed and risk-adjusted LOS reported with the predicted LOS measured using the Intensive Care Outcomes Model - Length-of-Stay (ICOMLOS).
1b.1. **Developer Rationale:** The length-of-stay of hospitalized patients has been demonstrated to be a contributor to cost. Nowhere is this more evident than in the intensive care unit, where the severity of illness requires costly technology to support such critically ill patients. The efficiency of ICU resource use along with overall quality of care can be measured as a means to compare performance between hospitals. Using the LOS measure, the hope is to identify modifiable factors enabling improvement in both ICU efficiency and effectiveness.

A 2007 analysis using the SAPS 3 database found that the presence of interprofessional rounds and an on-site emergency department were both factors that contributed to performance categorization of a hospital based on its risk-adjusted mortality and risk-adjusted LOS. A number of studies have also looked at intensivist staffing as a means of successfully reducing both ICU and hospital LOS. Adherence to process measures such as stress ulcer prophylaxis, deep vein thrombosis prophylaxis, appropriate use of transfusions, and appropriate sedation have also been reviewed in the literature in an effort to shorten ICU LOS. These are but a few of the potential structural features or care processes that may be influential in reducing the LOS outcome.

S.4. **Numerator Statement:** For all eligible patients admitted to the ICU, the time at discharge from ICU (either death or physical departure from the unit) minus the time of admission (first recorded vital sign on ICU flow sheet). The measure is risk-adjusted, please see S.18.
S.7. **Denominator Statement:** Total number of eligible patients who are discharged (including deaths and transfers)
S.10. **Denominator Exclusions:** <18 years of age at time of ICU admission, ICU readmission, <4 hours in ICU, primary admission due to trauma, burns, or immediately post-CABG, admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, transfers from another acute care hospital.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Paper Medical Records
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Jan 17, 2011 **Most Recent Endorsement Date:** Jan 17, 2011

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure is intended to be paired with risk-adjusted ICU mortality measure (NQF measure #0703) to avoid the creating the appearance that hospitals could or might lower LOS by allowing patients to die early.

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

- The developer attests there is new evidence since the last NQF review in 2011, but it appears only explanatory information has been provided.
- The developer provides the following rationale for this outcome measure: LOS of hospitalized patients, including in the ICU, has been demonstrated to be a contributor to cost. The efficiency of ICU resource use, along with overall quality of care, can be measured as a means to compare performance between hospitals.

Question for the Committee:

- *The rationale for the measure does not appear to have changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

- Performance scores are from data collected between 2010 and 2011.
- 224 hospitals contributed data, representing 69,510 patients and 236,374 ICU days, for an overall unadjusted mean LOS of 3.4 days. The standard deviation in LOS across hospitals was 0.8 days with an interquartile range of 2.8 to 3.9 days; the minimum LOS observed was 1.77 days, while the maximum was 6.69 days.
- The scores by decile are: 1.77 (0th percentile), 2.41, 2.76, 2.92, 3.11, 3.30, 3.52, 3.70, 3.99, 4.31, and 6.69 days (100th percentile).

Disparities

- The developer states disparities exist among different population groups, diagnosis and level-of care, but does not provide data based on use of the measure.
- The developer reports racial disparities also have been reported for African-Americans, whose adjusted ICU length of stay was significantly shorter than that of Whites. (Williams 1995)
- The last NQF Committee review noted cultural influences can affect LOS and that the developer should develop a means to address this and other disparities.

Questions for the Committee:

- *The most current gap information provided by the developer is from 2010-2011. Is the Committee aware of more current data on ICU LOS?*
- *Is there a gap in care that warrants a national performance measure?*
- *The developers note disparities exist in this area of healthcare, but do not provide current information. Are there other areas not noted by the developers that should be examined, i.e., racial groups other than African*

Americans and whites, age groups, within geographic locations (city vs. suburbs)?

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**Outcome measure based on paper medical records. Facility level measure.
Paired with risk-adjusted ICU mortality measure (NQF measure #0703) to avoid the creating the appearance that hospitals could or might lower LOS by allowing patients to die early.
Question for the Committee:

o The rationale for the measure does not appear to have changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence? YES--there is no new evidence; it is what it is.

**Reductions in ICU LOS will lead to reductions in cost of care.
**There is very little published data supporting this measure as a measure of quality. The developers mention association with cost, not quality.
**I agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence.

1b. Performance Gap

Comments:

**Questions for the Committee:

o The most current gap information provided by the developer is from 2010-2011. Is the Committee aware of more current data on ICU LOS? NO

o Is there a gap in care that warrants a national performance measure? YES

o The developers note disparities exist in this area of healthcare, but do not provide current information. Are there other areas not noted by the developers that should be examined, i.e., racial groups other than African Americans and whites, age groups, within geographic locations (city vs. suburbs)? PRIOR REVIEW MENTIONED CULTURAL? BUT UNCERTAIN HOW TO MEASURE OR ANALYZE THE SIGNIFICANCE?

**Compliance with this measure seems to entail a mean LOS that compares favorably with other centers who report data for the measure.
**The gap represented by a less than one day SD in LOS is very small and likely does not represent a size that warrants a measure. The developers do not provide adequate data on disparities.
**There appears to be performance gaps for ICU LOS across health systems that warrant a national performance measure.
Disparities have been documents for diagnostic groups as well as racial groups. I'm unaware of disparities for other groups.

1c. High Priority (previously referred to as High Impact)

Comments:

**Not applicable.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Paper medical records

Specifications:

- The developer attests there have been no change in specifications since the last NQF review.
- The numerator for this measure is: *For all eligible patients admitted to the ICU, the time at discharge from ICU (either death or physical departure from the unit) minus the time of admission (first recorded vital sign on ICU flow sheet).*
- The denominator of this measure: *Total number of eligible patients who are discharged (including deaths and transfers)*.
- The measure is risk-adjusted.
- The calculation algorithm is stated in S.18.

Questions for the Committee:

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The developer used data element-level reliability testing as published in *CHEST*, 2009; 136(1):89-101.
  - The data sample was comprised of 11,295 patients admitted to 35 California hospitals' ICUs from 2001-2004.
  - Inter-rater reliability was assessed. The developers reported agreement ranging from 91.5 to 98.8%, and weighted Kappa statistics ranging from 0.72-0.96.

Describe any updates to testing

- The developer indicates new reliability testing was conducted at the level of the measure score, but reports and justifies score-level testing for its related mortality measure as equivalent.
- The developer also conducted validity testing at the data element level by comparing hospital abstraction results to an auditor's. Per NQF guidance, separate reliability testing is not required when validity testing at the data element level is performed for all critical data elements. The developer states it assessed all individual risk model elements.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- For the reliability testing at the level of the performance score, the developer calculated the correlation between probability of death, not LOS, calculated on the data collected by the hospital's data collector and the probability of death calculated on the data collected by trained auditors (blinded to the data originally submitted by the hospital's data collector).
  - The developer states this is an appropriate type of testing because it provides a sense of how closely the scores agreed, but does not provide for any disagreement or indication of whether the hospital was reporting higher or lower severity than the auditor found.
  - The developer notes it was only publicly reporting the ICU mortality measure and not the LOS measure, so the audit assessed for over-reporting of risk only by calculating the predicted probability of death. Explicit calculations on the correlation of predicted LOS were not done, although the developer states the models predicting death and LOS used the exact same variables, though with different coefficients

on these variables. The developer posits over-reporting of risk factors for mortality (if present) should reflect over-reporting for LOS.

- For reliability testing at the data element level, the developer relies on validity testing at the data element level, as discussed in the following section.

Results of reliability testing

- For reliability testing at the score level (mortality, not LOS), the developer reported the correlation coefficient between the hospital's predicted probabilities of death and the auditor's predicted probabilities was 0.792. The developer state there was no clear pattern suggesting that hospitals over-reported risk factors—i.e., in some cases hospitals were over-reporting and in others they were under-reporting.

Guidance from the Reliability Algorithm (based on prior testing submission): 1→ 2 → 4 → 5→8 → 9 → 10 (highest eligible rating is MODERATE)

Question for the Committee:

- *Is the developer's reliability testing at the measure score level (new in this submission) appropriate (i.e., use of mortality as a proxy)? If so, per the Reliability Algorithm, the eligible ratings are HIGH, MODERATE, or LOW.*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

The developer provided the following information for the 2010-2011 endorsement maintenance review:

- 11,295 ICU patients from 35 California hospitals between 2001-2004
- 40% of the sample (n =4,611) was used for validation of the model.
- Model performance was assessed using:
  - A paired Student's t-test was used to compare mean observed ICU LOS to mean predicted ICU LOS for the entire validation population and for specific subgroups.
  - After dividing into deciles of predicted LOS, a paired Student's t-test and calibration curves were used to compare mean observed LOS to mean predicted LOS.
  - Coefficients of determination were calculated to measure the variance in LOS. Bivariate regression of the mean observed LOS against the mean predicted LOS was performed to assess the proportion of variation across hospitals explained by the model.
  - The assessment of the MPM III LOS model was compared to the performance of the ICU of each hospital by calculation of a SLOSR.
- The developer reported the following testing results:
  - Difference between the mean observed LOS and predicted LOS in the validation sample was 0.2 hours for MPM III LOS (p = 0.90).
  - The grouped hospital-level coefficient of determination for ICU LOS predictions was 0.279, indicating that 28% of ICU LOS variations were accounted for by MPM III LOS.
  - The SLOSRs of the MPM III LOS model ranged from 0.40 to 1.68.

Describe any updates to validity testing

- Additional empirical validity testing at the data element level has been conducted; previous testing was at the score level.
- Data were collected on ~70,000 patients per year. Audits were performed among hospitals participating in the California Hospital Assessment and Reporting Taskforce (CHART) program; 209 hospitals of all sizes in California could potentially be selected for random audits. The participating hospitals had 76% of all ICU patients in California. Almost all patients ≥18 years admitted to ICUs who stayed at least 4 hours were included. Hospitals included teaching and non-teaching institutions, rural (including Critical Access Hospitals) and urban institutions.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- The developer states it assessed agreement between trained auditors (the authoritative source) and hospital data collectors for all individual risk model elements.

Validity testing results:

The developer provides the following:

- Percent agreement between auditors and hospital data collectors across all individual risk model elements was 94%, with a range for specific risk variables from 85-97%.
- Kappas for specific variables ranged from 0.37-0.86, with most above 0.6 (substantial agreement or better). The developer does not specify Kappas for the individual variables.
- The developer did not provide sensitivity, specificity, positive predictive value, negative predictive value.
- The developer noted, based on Landis and Koch, only one variable fell below at least "moderate" agreement, the coma variable, with the auditors identifying more cases than the hospitals had coded.

Questions for the Committee:

- *The developer previously provided score-level testing and now provides data-element level testing. Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer noted the following exclusions: Trauma, burn, and coronary artery bypass surgery patients (whose outcomes are assessed using other methods—e.g., the Society of Thoracic Surgeons' bypass surgery mortality reports)

Questions for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

The developer noted the following:

- The LOS risk-adjustment model is based on the Intensive Care Outcomes Model - Length-of-Stay (ICOMLOS), with candidate interactions among variables and variable coefficients customized for the population of interest. Developer used the model that minimized data collection cost.
- Risk-adjustment variables (15) include: age, heart rate >=150, SBP <=90, chronic renal, acute renal, GIB, cardiac arrhythmia, intracranial mass effect, mechanical ventilation, received CPR, cancer, cerebrovascular incident, cirrhosis, coma, medical admission or status post non-elective surgery, zero factor status (no risk factors other than age), and full code status (no restrictions on therapies or interventions at the time of ICU admission).
- Discrimination and calibration statistics were calculated on the entire population.
- At the patient level, the model had an R^2 of 0.28, which the developer characterizes as good calibration. The developer notes that, at the hospital level, the correlation coefficient between this model and the APACHE model (which was the model requiring the most data collection and with 3 times the time required to collect all variables, but which had an R^2 of 0.42) was 0.89.
- The developer did not include SDS factors.
  - The developer performed an analysis on all 9,518 eligible patients admitted to the ICUs of 35 hospitals. For each patient, the developer had clinical risk factors and the following SDS variables: race, ethnicity, insurance status (including the categories of Medicaid and uninsured), and AHRQ Socioeconomic Status Index Score.
  - The developer reports none of the SDS variables were associated with mortality, even after accounting for differences in resuscitation preferences (Erickson S, et al. The Effect of Race and Ethnicity on Outcomes Among Patients in the Intensive Care Unit: A Comprehensive Study Involving Socioeconomic Status and Resuscitation Preferences. *Critical Care Medicine*, 2011; 39(3):429-35). The developer does not provide an analysis of SDS specific to LOS.
  - The developer speculates the lack of association between SDS and outcomes is because the entire course of care occurs within the hospital, where the mechanisms by which disparities occur (such as lack of access to care, food deserts, etc. may have less impact.)

Questions for the Committee:

- *Is the risk adjustment method appropriate?*
- *Do you agree with the developer's decision, based on its analysis, to exclude SDS factors from the risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer built a Bayesian model using hospital performance scores plus a random effect for each hospital. Hospitals were rated as different from expected if the coefficients on their variables were statistically significantly different from one.
- 95% confidence intervals on the point estimate of the outcome rate were calculated for each hospital using a Bayesian version of a hierarchical logistic regression model (implemented using WinBUGS). Using those confidence intervals, each individual hospital's performance was then compared to three benchmarks (the 10th, 50th, and 90th percentiles of performance among participating CHART hospitals). Hospitals were assigned to 1 of the 5 categories.
- Results from the most recent quarter (not identified more specifically by developer):
  - 46.8% of hospitals would have been labeled as having "Average" performance
  - 20.5% would have been labeled "Superior"
  - 10.7% would have been labeled "Above Average"
  - 19.5% would have been labeled "Below Average"
  - 19.5% would have been labeled "Below Average".
- The developer states the larger than expected number of performance outliers suggests there are real differences in hospital performance.

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- No data were missing; the submission website performs data quality control that ensures all data are included and in range.
- The developers also indicate missing data is not possible because the hospital cannot close out and submit an incomplete file.

Guidance from the Validity Algorithm: 1→ 2 → 3 → 6 → 7 → 8 (highest eligible rating is HIGH)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**2b1: Developer reports specifications consistent with evidence.

Question for the Committee:

o Are the specifications consistent with the evidence? YES

**Specifications are consistent with evidence

**Poor reliability leads to poor validity. In addition, the risk adjustment tool is not robust and does not correct for flaws mentioned above.

**Specifications are consistent with the evidence.

2a2. Reliability Testing

Comments:

**Updated empirical validity testing at the data element level has been conducted; previous testing was at the score level. Data were collected on ~70,000 patients per year. Audits were performed among hospitals participating in the California Hospital Assessment and Reporting Taskforce (CHART) program; 209 hospitals of all sizes in California could potentially be selected for random audits. The participating hospitals had 76% of all ICU patients in California.

Almost all patients ≥18 years admitted to ICUs who stayed at least 4 hours were included. Hospitals included teaching and non-teaching institutions, rural (including Critical Access Hospitals) and urban institutions.

The developer also conducted validity testing at the data element level by comparing hospital abstraction results to an auditor's.

Percent agreement between auditors and hospital data collectors across all individual risk model elements was 94%, with a range for specific risk variables from 85-97%. Kappas for specific variables ranged from 0.37-0.86, with most above 0.6 (substantial agreement or better). The developer does not specify Kappas for the individual variables. The developer did not provide sensitivity, specificity, positive predictive value, negative predictive value. The developer noted, based on Landis and Koch, only one variable fell below at least "moderate" agreement, the coma variable, with the auditors identifying more cases than the hospitals had coded.

Questions for the Committee:

o The developer previously provided score-level testing and now provides data-element level testing. Do the results demonstrate sufficient validity so that conclusions about quality can be made? YES

**This is not an "emeasure", it requires data collection from paper records; % agreement of auditors with hospital data collectors = 85-97%.

**Validity testing in California may not be representative of the nation.

**The data element level testing indicates this is a valid measure. However, the developers also state that the quality of the data reported depends on the quality of the data collectors and the developers recommend continuous training for data collectors.

2b2. Validity Testing

Comments:

**2b3: Exclusion:

Questions for the Committee:

o Are any patients or patient groups inappropriately excluded from the measure? NO

2b4: Risk Adjustment

Questions for the Committee:

o Is the risk adjustment method appropriate? YES

o Do you agree with the developer's decision, based on its analysis, to exclude SDS factors from the risk-

adjustment model? UNCERTAIN
2b5: Meaningful Difference
Question for the Committee:
o Does this measure identify meaningful differences in quality? YES
2b6: Not applicable
2b7: No missing data
Guidance from the Validity Algorithm: 1→ 2 → 3 → 6 → 7 → 8 (highest eligible rating is HIGH)
**I do not agree with excluding SDS factors from analysis.
**unclear why trauma, burns and cabg excluded
**Exclusions: appropriate
Risk Adjustment: appropriate
Meaningful Differences in quality appear to be able to be measured.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**The developer indicates new reliability testing was conducted at the level of the measure score, but reports and justifies score-level testing for its related mortality measure as equivalent.
Level of the performance score, calculated the correlation between probability of death, not LOS. While the developer states this is an appropriate type of testing because it provides a sense of how closely the scores agreed, but does not provide for any disagreement or indication of whether the hospital was reporting higher or lower severity than the auditor found.
For reliability testing at the score level (mortality, not LOS), the developer reported the correlation coefficient between the hospital's predicted probabilities of death and the auditor's predicted probabilities was 0.792.
Guidance from the Reliability Algorithm (based on prior testing submission): 1→ 2 → 4 → 5→8 → 9 → 10 (highest eligible rating is MODERATE)
Question for the Committee:
o Is the developer's reliability testing at the measure score level (new in this submission) appropriate (i.e., use of mortality as a proxy)? If so, per the Reliability Algorithm, the eligible ratings are HIGH, MODERATE, or LOW.
APPROPRIATE YES: MODERATE.
**Moderate rating
**Reliability testing appears to have been done for mortality and not LOS and then extrapolated without evidence that such extrapolation is justifiable.
**The use of mortality as a surrogate for ICU LOS appears to be reasonable, and I rate the reliability testing from the Reliability Algorithm to be MODERATE.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The measure requires chart abstraction.
- All data elements are not in defined fields in electronic sources. The developer is currently developing ICU eMeasures based on the Health Quality Measure Format (HQMF), to be completed in 2016.
- The developer reports each of 188 hospitals in California (full range of hospital types) submitted data on 400 patients/year (the first 100 consecutive patients in each quarter), and chart abstractors took ~11-15 minutes to collect data from each chart, which it considers minimal burden.
- There are no costs or licensing requirements.

Questions for the Committee:
o *Are the required data elements routinely generated and used during care delivery?*
o *Does the usefulness of the measure outweigh the data collection burden of manual chart abstraction?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments:

**Chart abstraction required. Developers developing ICU eMeasures. Takes 11-15 minutes to abstract data, considered minimal burden.

Questions for the Committee:

o Are the required data elements routinely generated and used during care delivery? GENERATED YES, USED IN CARE DELIVERY NO

o Does the usefulness of the measure outweigh the data collection burden of manual chart abstraction? YES

**As noted above, data for this measure will be obtained from paper records.

**ICU status is not tracked by most EMRs. ICU location is tracked but not all patients in the physical location are ICU status. Furthermore, a definition of ICU status is lacking

**Feasibility seems appropriate and will be improved with the implementation of the eMeasures reporting system.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- Until 2013, the measure were used for internal QI in California.
- In 2013, the developer began transforming this measure into an eMeasure for CMS consideration. Currently, a model using data from two hospital EMRs is in progress.

Improvement results

- Developer is unable to compute a trend in performance on this measure.

Unexpected findings (positive or negative) during implementation

- The developer reports no challenges/difficulties with implementation and no unexpected findings during implementation.

Potential harms

- The developer noted one potential unintended consequence: hospitals may seek to avoid high-risk patients (who, due to the severity of their illness, require longer ICU lengths-of-stay). The developer has not sought to assess this, but noted it could be monitored by evaluating changes in hospitals' risk-profiles over time.

Feedback

- No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

Questions for the Committee:

- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *There are no performance trends. Is this measure useful to further the goal of high-quality, efficient healthcare?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Not currently publicly reported nor accountability programs. Future use in accountability program planned. No improvement data available for analysis to be able to be reported. No unexpected findings during implementation. Single potential unintended consequence--hospitals may seek to avoid high-risk patients.

Questions for the Committee:

- Do the benefits of the measure outweigh any potential unintended consequences? UNCERTAIN
- There are no performance trends. Is this measure useful to further the goal of high-quality, efficient healthcare? YES, ESPECIALLY ONCE ELECTRONIC INFO AVAILABLE, HOWEVER WILL NEED TO BE BALANCED WITH RISK ADJUSTMENT.

Related or competing measures

- 0703 : Intensive Care: In-hospital mortality rate (harmonized)
- 0334: PICU Severity-adjusted Length of Stay (no info)

**With advancing trends for quality transparency, I am concerned that without valid adjustment there is a potential for referral centers to be misrepresented.

**The concern is that patients could be pushed out too early. The right/correct LOS per condition is not known. Furthermore, the correct LOS is complicated by severity of illness which is not presently well adjusted for. For instance, a patient with a single level laminectomy typically does not require ICU stay. However, that same patient with either a past history of a coagulopathy, say von Willebrand's disease or a past history of an unexplained episode of periop bleeding may very well benefit/require ICU stay. Such conditions are not captured by risk adjustment models that presently exist

**This measure is useful to further the goal of high-quality, efficient healthcare. Most consequences of its use are positive, and the negative consequences have been identified and monitoring for these consequences is recommended by the developer.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0703 : Intensive Care: In-hospital mortality rate
- 0334: PICU Severity-adjusted Length of Stay

Harmonization

- Developer indicates this measure and 0703 have been paired. The measure specifications are harmonized with NQF 0703. The previous Committee agreed pairing was necessary to balance potential unintended consequences of inappropriate reductions in LOS.
- Developer did not provide harmonization information for 0334.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0703

Measure Title: Intensive Care: In-hospital mortality rate

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: **Intensive Care: In-hospital mortality rate**

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Prevention of death is the reason why patients are admitted to the ICU. The critically ill are a uniquely vulnerable patient population, and given the high costs of their care, determining which hospitals have higher than expected rates of mortality following ICU admission is a meaningful measure of ICU performance.

The observed mortality variation among ICU patients from different hospitals has prompted in-depth study of possible structural or process contributors to this variation. The field of health services research has sought to identify any number of predictors, which include, but are not limited to, presence of an ICU medical director, presence of an intensivist, nurse-to-patient ratio, use of ventilator weaning protocols, and even ICU teamwork factors. For instance, hospitals associated with an ACGME residency have been shown to perform better than expected when comparing observed to expected rates of death. Similarly interdisciplinary clinical rounds and the presence of an on-site emergency department have also been shown to be statistically significant variables in mortality prediction. These are but a few of the factors that have been pursued in an effort to identify what makes certain hospitals perform better than others when mortality is the measured outcome.

There is variable evidence about the link between structure and outcome, although in general it is believed that, for ICUs that are not in teaching hospitals, having more intensivist coverage is associated with better outcomes.

However, there are many evidence-based processes of care that do improve outcomes, including ventilator bundles, sepsis management, and appropriate choices of drugs (e.g., beta blockers or the correct antibiotic) or interventions (e.g., PCI for MI) for heart disease, infections, etc.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

In the process of evaluating possible structural or procedural factors contributing to mortality variation, a number of studies have been able to identify certain variables that are less significant. For instance, hospital size (>300 vs <300 patients) was not found to impact standardized mortality ratio in one particular study of 35 California hospital ICUs. Another study similarly looking at patient volume found that higher ICU patient volumes were associated with lower mortality rates, but only in high-risk critically ill adults. In a worldwide sample using the SAPS 3 database, authors found that factors related to nursing or physician staffing had no impact on performance ratings. Such findings may be specific to the populations to which the variables were studied, but warrant further examination.

In addition, there are many condition-specific processes—such as administering beta blockers or performing percutaneous coronary intervention for patients with acute myocardial infarction, or adhering to the ventilator bundle for patients that require mechanical ventilation—that are associated with improving the outcomes of ICU patients.

Glance LG, Yue L et al. Impact of patient volume on the mortality rate of adult intensive care unit patients. Crit Care Med 2006;34(7):1925-34.

Kuzniewicz MW, Vasilevskis EE, Lane R et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008 Jun;133(6):1319-27.

Rothen HU, Stricker K, Einfalt E et al. Variability in outcome and resource use in intensive care units. Intensive Care Med 2007;33:1329-36.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report
Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form 0702_ICU_LOS_evidence_attachment_2015.docx
1b. Performance Gap Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating: • considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or • disparities in care across population groups. 1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)* The length-of-stay of hospitalized patients has been demonstrated to be a contributor to cost. Nowhere is this more evident than in the intensive care unit, where the severity of illness requires costly technology to support such critically ill patients. The efficiency of ICU resource use along with overall quality of care can be measured as a means to compare performance between hospitals. Using the LOS measure, the hope is to identify modifiable factors enabling improvement in both ICU efficiency and effectiveness. A 2007 analysis using the SAPS 3 database found that the presence of interprofessional rounds and an on-site emergency department were both factors that contributed to performance categorization of a hospital based on its risk-adjusted mortality and risk-adjusted LOS. A number of studies have also looked at intensivist staffing as a means of successfully reducing both ICU and

hospital LOS. Adherence to process measures such as stress ulcer prophylaxis, deep vein thrombosis prophylaxis, appropriate use of transfusions, and appropriate sedation have also been reviewed in the literature in an effort to shorten ICU LOS. These are but a few of the potential structural features or care processes that may be influential in reducing the LOS outcome.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Just as in-hospital mortality variation following ICU admission has been well-documented in the literature, so has variation in ICU LOS. One of the earlier publications on this subject (1993) in 42 ICUs among 40 volunteer hospitals reported a mean unadjusted length of ICU stay varying from 3.3 to 7.3 days, with 78% of the variation attributable to patient and selected institutional characteristics. More recent studies on different patient populations have since documented similar variation in ICU resource use and have made efforts to uncover reasons for this variability. Hospital geographic location has been interestingly found to be a significant contributor to ICU LOS in certain situations, though other structural and/or procedural variables are targets of further review.

The most recent performance scores for this measure are based on data collected between the third quarter of 2010 and the second quarter of 2011. 224 hospitals contributed data, representing 69,510 patients and 236,374 ICU days, for an overall unadjusted mean LOS of 3.4 days. The standard deviation in LOS across hospitals was 0.8 days with an interquartile range of 2.8 to 3.9 days; the minimum LOS observed was 1.77 days, while the maximum was 6.69 days. The scores by decile are: 1.77 (0th percentile), 2.41, 2.76, 2.92, 3.11, 3.30, 3.52, 3.70, 3.99, 4.31, and 6.69 days (100th percentile).

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Keenan SP, Dodek P, Martin C et al. Variation in length of intensive care unit stay after cardiac arrest: where you are is as important as who you are. Crit Care Med 2007;35:836-41.
Knaus WA, Wagner DP, et al. Variations in mortality and length of stay in intensive care units. Ann Int Med 1993;118:753-61.
Render ML, Kim M, Deddens J et al. Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure. Crit Care Med 2005;33(5): 930-9.
Rothen HU, Stricker K, Einfalt E et al. Variability in outcome and resource use in intensive care units. Intensive Care Med 2007;33:1329-36.
Vasilevskis EE, Kuzniewicz MW et al. Mortality Probability Model III and Simplified Acute Physiology Score II: assessing their value in predicting length of stay and comparison to APACHE IV. Chest 2009 Jul;136(1):89-101.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Disparities in ICU LOS do exist among different population groups. In an Italian study of patients with any of the following diagnoses - trauma, brain-trauma, brain-hemorrhage, stroke, acute-on-chronic-obstructive-pulmonary disease, lung-injury/acute respiratory distress syndrome, heart failure, and scheduled/unscheduled abdominal surgery - mean ICU variable-costs (and associated LOS) significantly differed with diagnosis and level-of-care. Other studies have documented higher costs per day in other diagnostic groups, such as septic patients or multiple trauma patients. In addition, racial disparity in LOS has even been reported for African-Americans, whose adjusted ICU length of stay was significantly shorter than that of whites.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Iapichino G, Radrizzani D et al. Effectiveness and efficiency of intensive care medicine: variable costs in different diagnostic groups. Acta Anaesthesiol Scand 2004 Aug;48(7):820-6.
Moerer O, Plock E, Mgbor U et al. A German national prevalence study on the cost of intensive care: an evaluation from 51 intensive care units. Crit Care 2007;11(3):R69.
Rossi C, Simini B, Brazzi L et al. Variable costs of ICU patients: a multicenter prospective study. Intensive Care Med 2006 Apr;32(4):545-52.
Williams JF, Zimmerman JE et al. African-American and white patients admitted to the intensive care unit: is there a difference in therapy and outcome? Crit Care Med 1995 Apr;23(4):626-36.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

High resource use, Affects large numbers, Frequently performed procedure, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

ICU resource use is viewed as a key indicator in assessing ICU performance. However, cost data are rather difficult to collect. ICU LOS, however, has become a surrogate for cost due to its relatively easy definability and measurability. One study even reported that length of stay statistically explains approximately 85 to 90% of interpatient variation in hospital costs. By 2005, critical care costs in the US were estimated to be $81.7 billion accounting for 13.4% of hospital costs, 4.1% of the national health expenditures and 0.66% of the gross domestic product. With mean estimated ICU costs estimated to be greater than $30,000 (when patients are mechanically ventilated) and initial ICU days found to be four times as costly as initial non-ICU hospital days, reductions in ICU LOS are viewed as a potential target for cost-cutting efforts.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Dasta JF, McLaughlin TP et al. Daily cost of an intensive care unit day: the contribution of mechanical ventilation. Crit Care Med 2005 Jun;33(6):1266-71.

Halpern NA. Can the costs of critical care be controlled? Curr Opin Crit Care 2009 Oct 9. [Epub ahead of print]

Rapoport JTD, Zhao Y, Lemeshow S. Length of stay data as a guide to hospital economic performance for ICU patients. Medical Care 2003;41:386-97.

Rosenberg AL, Zimmerman JE, Alzola C et al. Intensive care unit length of stay: recent changes and future challenges. Crit Care Med 2000 Oct 28(10):3465-73.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to*

general information.)
http://healthpolicy.ucsf.edu/sites/healthpolicy.ucsf.edu/files/documents/ICOM_Tool.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** ICU Outcomes Data Dictionary.pdf

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
No changes have been made.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
For all eligible patients admitted to the ICU, the time at discharge from ICU (either death or physical departure from the unit) minus the time of admission (first recorded vital sign on ICU flow sheet). The measure is risk-adjusted, please see S.18.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Not-applicable; anyone with an ICU admission meeting eligibility criteria below is in the numerator.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Eligible patients include those with an ICU stay of at least 4 hours and =18 years of age whose primary reason for admission does not include trauma, burns, or immediately post-coronary artery bypass graft surgery (CABG), as these patient groups are known to require unique risk-adjustment. Only index (initial) ICU admissions are recorded given that patient characteristics of readmissions are known to differ.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Total number of eligible patients who are discharged (including deaths and transfers)

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Eligible patients include those with an ICU stay of at least 4 hours and =18 years of age whose primary reason for admission does not include trauma, burns, or immediately post-coronary artery bypass graft surgery (CABG), as these patient groups are known to require unique risk-adjustment. Only index (initial) ICU admissions are recorded given that patient characteristics of readmissions are known to differ.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
<18 years of age at time of ICU admission, ICU readmission, <4 hours in ICU, primary admission due to trauma, burns, or immediately post-CABG, admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, transfers from another acute care hospital.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as*

definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)
<18 years of age at time of ICU admission (with time of ICU admission abstracted preferably from ICU vital signs flowsheet), ICU readmission (i.e. not the patient´s first ICU admission during the current hospitalization), <4 hours in ICU, primary admission due to trauma, burns, or immediately post-CABG, admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, patient transfers from another acute care hospital (i.e. patients whose physical site immediately prior to the index ICU admission was an acute care unit at an outside hospital).

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not-applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Risk-adjustment variables include: age, heart rate >=150, SBP <=90, chronic renal, acute renal, GIB, cardiac arrhythmia, intracranial mass effect, mechanical ventilation, received CPR, cancer, cerebrovascular incident, cirrhosis, coma, medical admission or status post nonelective surgery, zero factor status (no risk factors other than age), and full code status (no restrictions on therapies or interventions at the time of ICU admission). The LOS risk-adjustment model is based on the Intensive Care Outcomes Model - Length-of-Stay (ICOMLOS) with candidate interactions among variables and variable coefficients customized for the population of interest.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Model specifications are available online at:
http://healthpolicy.ucsf.edu/sites/healthpolicy.ucsf.edu/files/documents/ICU_Outcomes_Models_S9.pdf

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 hospital´s mean observed ICU LOS and and mean risk-adjusted LOS are calculated using the abstracted data. For each hospital, the model produces a median and 95% confidence interval for the standardized LOS ratio (SLOSR), which is the mean observed LOS divided by the mean predicted LOS.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF a PRO-PM, identify whether (and how) proxy responses are allowed.
the first 100 consecutive eligible patients per quarter

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
In our implementation of this, missing data was not possible (because the hospital cannot close out and submit an incomplete file). If other implementation approaches were taken, for any missing variables we would recommend an assumption that the variable for which a value is missing is in the normal range (that is, the blood pressure was normal or the patient did NOT have a diagnosis risk factor such as cancer). This approach incentivizes the hospital to provide complete and accurate data.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
ICU Outcomes Data Collection Instrument

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0702_ICU_LOS_Testing_attachment_2015_1214_REV.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0702
Measure Title: Intensive Care Unit (ICU) Length-of-Stay (LOS)
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☒ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Disclaimer: the older information in blue below may not completely match with the layout of the new form.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From:	Measure Tested with Data From:

(*must be consistent with data sources entered in S.23*)	
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

ORIGINAL ANSWER from ORIGINAL QUESTION 2a2.1 Data /Sample:

11,295 ICU patients from 35 California hospitals between 2001-2004.

ORIGINAL ANSWER from ORIGINAL QUESTION 2b4.1 Data /Sample and QUESTION 2b2.Data /Sample:

6,684 patients were used in the development sample in order to estimate coefficients for the MPM III LOS model, and 40% of the sample (n =4,611) was used for validation of the model.

Original reliability testing methods and results are taken from *CHEST*, 2009; 136(1):89-101, in which the data sample was comprised of 11,295 patients admitted to 35 California hospitals' ICUs from 2001-2004. The models were then updated each year in our work with the California Hospital Assessment and Reporting Taskforce. In this program data were collected on ~70,000 patients per year. The reliability testing described below reflects audits performed among hospitals participating in the CHART program.

1.3. What are the dates of the data used in testing? October 2010-September 2011

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*) **(new question, not in previous form)**

209 hospitals of all sizes in California were in the pool to potentially be selected for random audits (more details in 1.7 below). Hospitals included teaching and non-teaching institutions, rural (including Critical Access Hospitals) and urban institutions.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*) **(new question, not in previous form)**

Almost all patients ≥18yo admitted to ICUs who stayed at least 4 hours were included. Exclusion criteria were: 1) having a primary reason for admission of trauma, burns, and coronary artery bypass graft (CABG) surgery patients (because other performance measures exist for these populations), 2) admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, and 3) transfers from another acute care hospital. Only index (initial) ICU admissions are recorded given that patient characteristics of readmissions are known to differ. Collectively, the participating hospitals had 76% of all ICU patients in California. Each hospital was asked to submit data on 400 patients a year (the first 100 consecutive patients in each quarter).

Discrimination and calibration statistics for the risk adjustment model were calculated on the entire population. Auditing of the reliability of critical data elements involved having a second reviewer re-abstract all charts and was performed on 986 patients from (34 each from 29 hospitals selected randomly from each performance group with intentional over-sampling of hospitals rated as performing better or worse than expected in terms of ICU mortality, since this was the outcome expected most likely to induce gaming—if any gaming were to occur). This approach gave us 60% power to detect hospitals that were truly misclassified in terms of ICU mortality performance.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate). **(new question, not in previous form)**

SDS variables were not available and were not used. See section 2b4.4b for evidence that race, ethnicity, insurance status and AHRQ Socioeconomic Status Index Score variables were not associated with ICU outcomes.

2a2. RELIABILITY TESTING

***<u>Note</u>**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data were reabstracted by auditors on a 5% random sample of patients. Kappa statistics were calculated for inter-rater variability between the data abstractor and the auditor. The auditors were clinical nurses who were trained by the authors and completed extensive sample chart abstraction.

See section 2b2 for validity testing of individual data elements.

For the reliability testing of the performance score, we calculated the correlation between probability of death calculated on the data collected by the hospital's data collector and the probability of death calculated on the data collected by trained auditors (blinded to the data originally submitted by the hospital's data collector). This provides a sense of how closely the scores agreed, but does not provide (for any disagreement), any indication of whether the hospital was reporting higher or lower severity than the auditor found.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

For physiologic variables of the MPM III LOS model, inter-rater reliability was excellent, with agreement ranging from 91.5 to 98.8%, and weighted kappa statistics ranging from 0.72-0.96.

See section 2b2 for validity testing of individual data elements.

Because we were publicly reporting the ICU mortality measure but not the LOS measure, in our audit we assessed for over-reporting of risk only by calculating the predicted probability of death. However, although we did not explicitly calculate the correlation of predicted LOS, since the models predicting death and LOS used the exact same variables (though with different coefficients on these variables, of course), over-reporting of risk factors for mortality (if present) should reflect over-reporting for LOS. Therefore, we now share the data for mortality:

The correlation coefficient between the hospital's predicted probabilities of death and the auditor's predicted probabilities was 0.792. There was no clear pattern suggesting that the hospitals over-reported risk factors (that is, in some cases, hospitals were over-reporting, in others, they were under-reporting).

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
The data collection for this measure requires training and mistakes can be made both in favor of and against the reporting hospitals. Periodic audits and training of data collection staff are recommended. On the whole, though, the errors did not seem to reflect gaming of the system (though this was only tested in a context of public reporting, not a payment). Thus, we conclude that, like all other measures, it is likely that attention needs to be paid to data quality.

2b2. VALIDITY TESTING

ORIGNIAL 2b2.1 ANSWER MOVED TO QUESTION 1.2 OF THIS FORM:

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**
☐ **Empirical validity testing**
☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
In order to assess model performance in the validation sample, multiple methods were used:

1. A paired Student's t-test was used to compare mean observed ICU LOS to mean predicted ICU LOS for the entire validation population and for specific subgroups.

2. After dividing into deciles of predicted LOS, a paired Student's t-test and calibration curves were used to compare mean observed LOS to mean predicted LOS.

3. Coefficients of determination were calculated to measure the variance in LOS. Bivariate regression of the mean observed LOS against the mean predicted LOS was performed to assess the proportion of variation across hospitals explained by the model.

4. The assessment of the MPM III LOS model was compared to the performance of the ICU of each hospital by calculation of a SLOSR.

For the validity tests conducted, we used the kappa statistic, which expresses how much better or worse the agreement between two observers was compared to chance alone. Testing was conducted on 986 patients from (34 each from 29 hospitals selected to over-sample hospitals rated as performing better or worse than expected, with 60% power to detect hospitals that were truly misclassified).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Difference between the mean observed LOS and predicted LOS in the validation sample was 0.2 hours for MPM III LOS (p = 0.90). MPM III LOS had a single age stratum with significant differences between observed and predicted LOS. However, it accurately predicted ICU LOS for medical and elective surgical patients. The MPM III LOS model's calibration curve demonstrated excellent fit across deciles of predicted ICU LOS. The grouped hospital-level coefficient of determination for ICU LOS predictions was 0.279, indicating that 28% of ICU LOS variations were accounted for by MPM III LOS. The SLOSRs of the MPM III LOS model ranged from 0.40 to 1.68.

Percent agreement between auditors and hospital data collectors across all individual risk model elements was 94%, with a range for specific risk variables from 85-97%. Kappas for specific variables ranged from 0.37-0.86, with most above 0.6 (substantial agreement or better). The only variable below Koch and Landis' criteria for at least moderate agreement was the coma variable (kappa 0.37), with the auditors identifying more cases than the hospitals had coded.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
The data collection for this measure requires training and mistakes can be made both in favor of and against the reporting hospitals. Periodic audits and training of data collection staff are recommended. On the whole, though, the errors did not seem to reflect gaming of the system (though this was only tested in a context of public reporting, not a payment). Thus, we conclude that, like all other measures, it is likely that attention needs to be paid to data quality.

{This is Landis and Koch's interpretation scale for kappas: "The kappa statistic has a maximum value of 1 when

agreement is perfect, zero when agreement is no better than chance. The nomenclature of Landis and Koch is usually used to interpret the degree of agreement: kappa <0.21, slight agreement; 0.21-0.40, fair agreement; 0.41- 0.60, moderate agreement; 0.61-0.80, substantial agreement; and 0.81-1.00, almost perfect agreement." [Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174]}

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

Trauma, burn, and coronary artery bypass surgery patients (whose outcomes are assessed using other methods—e.g., the Society of Thoracic Surgeons' bypass surgery mortality reports), were excluded.

ORIGINAL 2b3.1 QUESTION: "Data/Sample for analysis of exclusions":

ORIGINAL ANSWER: Not-applicable

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Because these exclusions were pre-specified before the data collection began, no testing of the exclusions has been performed

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not-applicable

Not applicable.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not-applicable

Not applicable.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

ORIGINAL 2b4.1 ANSWER MOVED TO QUESTION 1.2 OF THIS FORM:

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 15 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

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

ORIGINAL ANSWER: Using all the variables in the original MPM III mortality model, mixed-effects, multilevel modeling was used to generate an ICU LOS prediction model based on the MPM III. The LOS was calculated in days to the second significant digit and truncated at 30 days to minimize the impact of outliers (as previous investigators have done).

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

ORIGINAL 2b4.3 QUESTION: "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): "

ORIGINAL ANSWER: Not-applicable

[No answer needed; measure is risk adjusted]

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

In building this model, we first considered all variables in all previously published ICU outcomes risk adjustment models (see *CHEST* 2008 Jun;133(6):1319-27, and *CHEST*, 2009; 136(1):89-101). With stakeholder input, we compared the models' predictive accuracy (discrimination and calibration) vs the cost of data collection (using chart abstractor time and motion studies). As it turned out, hospital performance ratings varied little among the different risk adjustment models, but the cost of data collection varied substantially. Therefore, the stakeholders chose the model that minimized data collection cost.

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

ORIGINAL ANSWER: Not-applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors? (new question, not in previous form)

At the patient level, the model had an R^2 of 0.28 with good calibration. At the hospital level, the correlation coefficient between this model and the APACHE model (which was the model requiring the most data collection and with 3 times the time required to collect all variables, but which had an R^2 of 0.42) was 0.89.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects) ? (new question, not in previous form)

We do not include SDS factors in our ICU models. In the development work for these models, we performed an analysis on all 9,518 eligible patients admitted to the ICUs of 35 hospitals (using the same eligibility criteria as are proposed for this measure). For each patient, we had, in addition to the clinical risk factors, the following

SDS variables: race, ethnicity, insurance status (including the caterogies of Medicaid and uninsured), and AHRQ Socioeconomic Status Index Score. *None of the SDS variables were associated with mortality* (Erickson, S, Vasilevskis, EE, et al. The Effect of Race and Ethnicity on Outcomes Among Patients in the Intensive Care Unit: A Comprehensive Study Involving Socioeconomic Status and Resuscitation Preferences. *Critical Care Medicine*, 2011; 39(3):429-35). This was true even after accounting for differences in resuscitation preferences. We suspect that the lack of association between SDS and outcomes seen here is because the entire course of care occurs within the hospital, where the mechanisms by which disparities occur (such as lack of access to care, food deserts, etc) may have less impact.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*) **(new question, not in previous form)**

Discrimination was assessed using R^2, calibration using calibration curves.

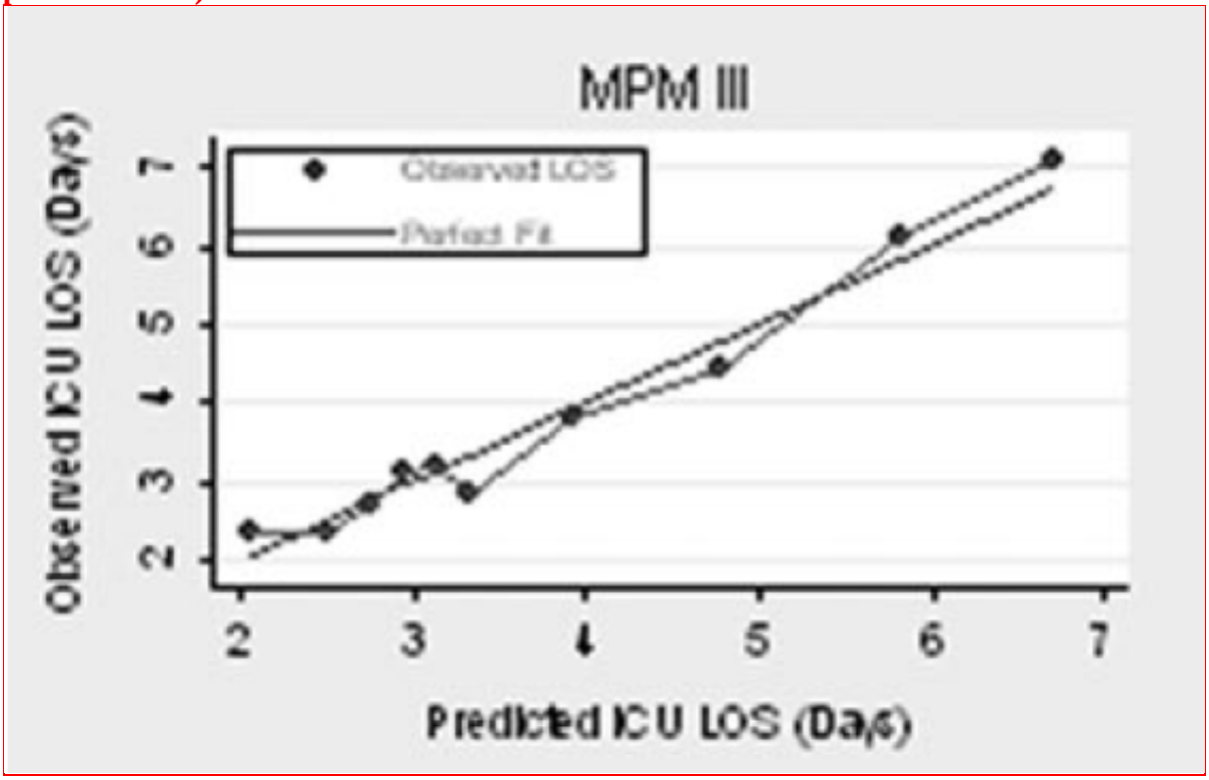
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**: (new question, not in previous form)**

R^2 of 0.28

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): **(new question, not in previous form)**

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: **(new question, not in previous form)**



2b4.9. Results of Risk Stratification Analysis: **(new question, not in previous form)**

Not applicable.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*) **(new question, not in previous form)**

The model demonstrates adequate discrimination and calibration.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

In order to compare predictions of the models for hospital-level performance, a plot of LOS prediction model-specific SLOSRs for each hospital with at least 100 admissions was generated.

A Bayesian model was built using hospital performance scores plus a random effect for each hospital. Hospitals were rated as different from expected if the coefficients on their variables were statistically significantly different from one.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

There were similar ranges among the SLOSRs for each model as follows:

recalibrated APACHE IV LOS 0.47-1.60
MPM III LOS 0.40 - 1.68
SAPS II LOS 0.38-1.69
The intraclass correlations of the SLOSRs between each pair of models was high:
recalibrated APACHE IV LOS and MPM III LOS r = 0.89 (95% CI, 0.74-0.96)
recalibrated APACHE IV LOS and SAPS II LOS r = 0.85 (95% CI, 0.70-0.93)
MPM III LOS and SAPS II LOS r = 0.96 (95% CI, 0.92-0.98)

Our public reporting program, CHART, has not yet decided to publicly report ICU LOS. However, we can assess what would happen in its program. CHART uses five levels of ratings: “Superior”, “Above Average”, “Average”, “Below Average”, and “Poor”. As a first step, 95% confidence intervals on the point estimate of the outcome rate were calculated for each hospital using a Bayesian version of a hierarchical logistic regression model (implemented using WinBUGS). Using those confidence intervals, each individual hospital's

performance was then compared to three benchmarks (the 10th, 50th, and 90th percentiles of performance among participating CHART hospitals). Hospitals were assigned to 1 of the 5 categories using the table below.

CHART Methodology Table

1. CHART Group Calculation

a) Using statistical model with hospital rates as random effects, using the program WinBUGS, hospital effect point estimates and 95% confidence intervals were calculated.

b) A low, middle and high reference value was obtained, as the 10th, 50th, and 90th weighted percentile rate. By definition, 10% of all patients were in hospitals whose rate was less than or equal to the 10th weighted percentile. Similarly for the 50th and 90th weighted percentiles.

2. Mortality Performance Group calculation

A hospital's confidence interval was compared to the low, middle, and high reference values to determine the CHART Group as follows:

"L" represents the location of the hospital's lower confidence limit
"U" represents the location of the hospital's upper confidence limit

2. Mortality Performance Group	2. Mortality Performance Group	2. Mortality Performance Group	2. Mortality Performance Group	2. Mortality Performance Group
Below Low Reference	Between the Low and Middle	Between Middle and Top	Above Top Reference	CHART Performance Group
L & U				1
L	U			1
	L & U			2
L		U		2
	L	U		3
		L & U		3
	L		U	4
		L	U	4
			L & U	5

Mortality Reference Values		
Low	Middle	High
6.87%	11.47%	15.75%

3. CHART Performance Group Scoring Key

1 - Superior
2 - Above Average
3 - Average
4 - Below Average
5 - Poor

New ratings were generated every quarter. In the most recent quarter in which ICU LOS was measured, 46.8% of hospitals would have been labeled as having "Average" performance. Of the remaining hospitals, 20.5% would have been labeled "Superior", 10.7% would have been labeled "Above Average", 19.5% would have been labeled "Below Average", and 0.2% would have been labeled "Poor".

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e.*, *what do the results mean in terms of statistical and meaningful differences?*)
We believe that the larger than expected number of performance outliers suggests that there were real differences in hospital performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than***

one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not-applicable

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not-applicable

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

(new question, not in previous form)
Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

(new question, not in previous form)

All data were submitted through a secure website that performs data quality control during the submission process, so no data were missing (each chart cannot be closed with missing or out of range data).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

(new question, not in previous form)

Not applicable.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

(new question, not in previous form)

Not applicable.

3. Feasibility
Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order). 3a.1. **Data Elements Generated as Byproduct of Care Processes.** Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry) If other:
3b. Electronic Sources The required data elements are available in electronic health records or other electronic sources. If the required data are not in

electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**
We are currently in the process of developing ICU eMeasures based on the Health Quality Measure Format (HQMF). We anticipate that there will be two versions of this mortality eMeasure: one to compute the raw (unadjusted) mortality rate and another to extract the data elements to compute risk-adjusted rates. The eMeasure will also incorporate ICU risk adjustment methodology that is currently under review for publication. We anticipate that development of this eMeasure will be completed in 2016.

The HQMF for this eMeasure will be developed using the Measure Authoring Tool based on the Quality Data Model (QDM). In addition, we anticipate that we will also be able to define Category I-III Quality Reporting Document Architecture (QRDA) documents with the eMeasure HQMF.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
In 188 hospitals in California (from small rural hospitals to the largest teaching hospitals), we have successfully collected this data. The average time per chart for an experienced data collector is 11-15 minutes. We collected data on 100 patients per quarter to minimize the data collection burden while still getting sufficient sample size to get precise estimates of hospital performance. However, an alternative target sample size could easily be chosen by users.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees or licensing requirements.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Not applicable.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

These measures were used for internal QI in California until 2013. At that time, the decision was made to focus on measures required by CMS. Therefore, we are currently working to transform this model into an e-measure for CMS consideration.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

We intend to transform to e-measures and have made substantial progress in that direction. We have been able to implement a model that uses data from 2 hospital EMRs and has superior predictive performance to the existing model. A paper describing this is under review. However, we have not yet transformed this model into HQMF. We expect this to be completed in 2016.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

We do not have the necessary data to be able to compute a trend in performance on this measure. However, on 0703 (ICU mortality) a trend of 13.85% to 11.67% was observed from 2007 to 2011, representing a decline in mortality of -2.18%. 69,483 patients and 187 hospitals were represented in these data, which was collected as a part of the California Hospital Assessment and Reporting Taskforce (CHART) project.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

This measure was never publicly reported and was not a focus of quality improvement. However, interest in efficiency measures has increased significantly recently, and the progress made on the mortality measure derived from the same variables demonstrates that organizations can use this information to improve.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
The potential unintended consequence is that hospitals may seek to avoid high-risk patients (who, due to the severity of their illness, require longer ICU lengths-of-stay). One could monitor this behavior by evaluating changes in hospitals' risk-profiles over time.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0703 : Intensive Care: In-hospital mortality rate

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses 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.**
This measure is completely harmonized with measure 0703 Intensive Care: In-hospital mortality rate.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Philip R. Lee Institute for Health Policy Studies
Co.2 **Point of Contact:** R. Adams, Dudley , adams.dudley@ucsf.edu, 415-476-8617-
Co.3 **Measure Developer if different from Measure Steward:** Philip R. Lee Institute for Health Policy Studies
Co.4 **Point of Contact:** R. Adams, Dudley , adams.dudley@ucsf.edu, 415-476-8617-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
not-applicable

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 07, 2009
Ad.4 **What is your frequency for review/update of this measure?** annually when in use
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.6 **Copyright statement:** is in the public domain
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0703
De.2. **Measure Title:** Intensive Care: In-hospital mortality rate
Co.1.1. **Measure Steward:** Philip R. Lee Institute for Health Policy Studies
De.3. **Brief Description of Measure:** For all adult patients admitted to the intensive care unit (ICU), the percentage of patients whose hospital outcome is death; both observed and risk-adjusted mortality rates are reported with predicted rates based on the Intensive Care Outcomes Model - Mortality (ICOMmort).
1b.1. **Developer Rationale:** Prevention of death is the reason why patients are admitted to the ICU. The critically ill are a uniquely vulnerable patient population, and given the high costs of their care, determining which hospitals have higher than expected rates of mortality following ICU admission is a meaningful measure of ICU performance.

S.4. **Numerator Statement:** Total number of eligible patients whose hospital outcome is death. The measure is risk-adjusted, please see S.18.
S.7. **Denominator Statement:** Total number of eligible patients who are discharged (including deaths and transfers out to other hospitals).
S.10. **Denominator Exclusions:** <18 years of age at time of ICU admission, ICU readmission, <4 hours in ICU, primary admission due to trauma, burns, or immediately post-CABG, admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, transfers from another acute care hospital.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Paper Medical Records
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Jan 17, 2011 **Most Recent Endorsement Date:** Jan 17, 2011

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure does not require pairing. However, it is recommended that, if measure 702 (Intensive Care: ICU Length of Stay) is used, that measure be paired with this measure to address concerns expressed during the original review that hospitals might lower length of stay by allowing rapid mortality.

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the

relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- The developer states there is new evidence since the last review in 2011, but NQF staff's review of the previous submission does not indicate updated literature.
- The developer provides the following rationale for this outcome measure: Prevention of death is the reason why patients are admitted to the ICU. The critically ill are a uniquely vulnerable patient population, and given the high costs of their care, determining which hospitals have higher than expected rates of mortality following ICU admission is a meaningful measure of ICU performance."
- The developer's previous submission provides examples of where structural or procedural factors may be less significant in improving outcomes for ICU patients, but it also notes many condition-specific processes are associated with improving the outcomes of ICU patients.

Questions for the Committee:

- *The rationale for the measure does not appear to have changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developers provide the following:

- Performance scores are from data collected between 2010 and 2011.
- 226 hospitals contributed data, representing 69,483 patients and 8,112 deaths, for an overall unadjusted mean mortality rate of 11.67%.
- The standard deviation in this rate across hospitals was 4.17%, with an interquartile range of 8.56% to 13.68%; the minimum rate observed was 1.31%, while the maximum was 28.16%.
- The scores by decile are: 1.31% (0th percentile), 6.74%, 8.06%, 9.2%, 10.29%, 11.4%, 12.59%, 13.19%, 14.31%, 16.12%, and 28.16% (100th percentile).
- Using a baseline sample in 2007, the observed unadjusted mortality rate was 13.85%. When the developer compares this to the more recent sample (mortality rate =11.67%), this represents a decline in mortality of 2.18%.

Disparities

- The developer notes the literature documents disparities in mortality outcomes following admission to the ICU among various populations:
  - disease-specific racial variation have been noted among blacks
  - the elderly, particularly older women fare worse than men
  - insurance status

Questions for the Committee:

- *The most current gap information provided by the developer is from 2010-2011. Is the Committee aware of more current data on ICU mortality?*
- *Is there a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**Outcome measure; paper medical records; facility level
Paired with 702

New evidence noted but not found
Questions for the Committee:
o The rationale for the measure does not appear to have changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence? YES
**Prevention of mortality is an essential goal of ICU care. This is an important quality measure.
**Evidence supports
**I agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence.
**Rationale hasn't changed since last endorsement, no need to repeat evidence

1b. Performance Gap
Comments:
**Questions for the Committee:
o The most current gap information provided by the developer is from 2010-2011. Is the Committee aware of more current data on ICU mortality? NO
o Is there a gap in care that warrants a national performance measure? YES
**Compliance with the measure entails a rate of mortality that compares favorably with other centers who are reporting: currently the overall adjusted rate if 11/67% +/- 4.1%.
**Gap exists
**I am not aware of more recent date.
The data presented indicate there is a gap in care that warrants a national performance measure.
**No newer data on ICU mortality. Aware that CERNER is working on algorithm for a new APACHE (V).
There is the need for a national performance measure as ICU care demand is expected to increase and cost mat approach 47.5% of hospital billings based on AHRQ's HCUP 2014 report # 185 (using 2011 data)

1c. High Priority (previously referred to as High Impact)
Comments:
**Not applicable.
**Yes. AUCs in a very acceptable range

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Paper medical records

Specifications:

- The developer attests there has been no change in measure specifications since the last NQF review.
- The numerator for this measure is: *Total number of eligible patients whose hospital outcome is death.*
- The denominator for this measure is: *Total number of eligible patients who are discharged (including deaths and transfers out to other hospitals).*
- The measure is risk-adjusted.
- The calculation algorithm is stated in S.18.

Questions for the Committee:

- o *Are all the data elements clearly defined?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The developer used testing as published in *CHEST*, 2009; 136(1):89-101.
  - The data sample was comprised of 11,300 patients admitted to 35 California hospitals' ICUs from 2001-2004.
  - Inter-rater reliability was assessed between the data abstractor and an auditor. The developer reported agreement ranging from 91.5 to 98.8%, and weighted kappa statistics ranging from 0.72-0.96.
- Although reported in the previous submission as reliability testing, NQF considers this to be validity testing at the data-element level (the auditor as the authoritative source). Per NQF guidance, separate reliability testing is not required when validity testing at the data element level is performed for all critical data elements.

Describe any updates to testing

- The developer indicates new, updated reliability testing at the level of the measure score.
- The developer also provides updated validity testing at the data element level by comparing hospital abstraction results to an auditor's. Per NQF guidance, separate reliability testing is not required when validity testing at the data element level is performed for all critical data elements. The developer states all critical data elements were audited.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- For the reliability testing of the performance score, the developer calculated the correlation between probability of death calculated on the data collected by the hospital's data collector and the probability of death calculated on the data collected by trained auditors (blinded to the data originally submitted by the hospital's data collector).
- The developer indicates this is an appropriate type of testing because it provides a sense of how closely the scores agreed, but does not provide for any disagreement or indication of whether the hospital was reporting higher or lower severity than the auditor found.

Results of reliability testing

- For reliability testing at the score level, the developer reported "the correlation coefficient between the hospital's predicted probabilities of death and the auditor's predicted probabilities was 0.792." The developer states there was no clear pattern suggesting hospitals over-reported risk factors—i.e., in some cases, hospitals were over-reporting, in others, they were under-reporting).
- The developer does not provide additional statistical analyses.

Guidance from the Reliability Algorithm: 1→ 2 → 3 → 6→7 → 8 (eligible for HIGH rating)

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the

evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Questions for the Committee:

- *None*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

The developer provided the following information for the 2010-2011 endorsement maintenance review:

- The developer used 50,307 (40.3%) of the patients to validate the model.
- The developer tested the model performance by using the area under the receiver operating characteristics curve (ROC) >=0.75 as a measure of discrimination and the Hosmer-Lemeshow goodness-of-fit statistic as a measure of calibration.
- The developer reports the following testing results:
  - area under the ROC curve was 0.823 (95% CI 0.818-0.828)
  - the Hosmer-Lemeshow statistic was 11.62 (p = 0.31),
  - the standardized mortality ratio was 1.018 (95% CI 0.996-1.040).

Describe any updates to validity testing

- Updated validity testing at the data element level has been conducted since the last endorsement review by comparing hospital abstraction results to an auditor's (the authoritative source).
- Data were collected on ~70,000 patients per year. Audits were performed among hospitals participating in the California Hospital Assessment and Reporting Taskforce (CHART) program; 187 hospitals of all sizes in California could potentially be selected for random audits. The participating hospitals had 76% of all ICU patients in California. Almost all patients ≥18 years admitted to ICUs who stayed at least 4 hours were included. Hospitals included teaching and non-teaching institutions, rural (including Critical Access Hospitals) and urban institutions.
- The data were tested between October 2010 and September 2011. The audit assessed 986 patients from 29 hospitals. Selected hospitals were rated as performing better or worse than expected, with 60% power to detect hospitals that were truly misclassified.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- For the new validity testing at the data element-level, the developer assessed agreement between trained auditors (the authoritative source) and hospital data collectors for all individual risk model elements.

Validity testing results:

The developer provides the following:

- Percent agreement between auditors and hospital data collectors across all individual risk model elements was 94%, with a range for specific risk variables from 85-97%.
- Kappas for specific variables ranged from 0.37-0.86, with most above 0.6 (substantial agreement or better). The developer does not specify Kappas for the individual variables.
- The developer did not provide sensitivity, specificity, positive predictive value, negative predictive value.

- The developer noted, based on Landis and Koch, only one variable fell below at least "moderate" agreement, the coma variable, with the auditors identifying more cases than the hospitals had coded.
- The developer states that while testing was only conducted in a context of public reporting, not payment, gaming of the system was not detected.

Questions for the Committee:

- *The developer previously provided score-level testing and now provides data-element level testing. Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developers noted the following exclusions: Trauma, burn, and coronary artery bypass surgery patients (whose outcomes are assessed using other methods—e.g., the Society of Thoracic Surgeons' bypass surgery mortality reports)

Questions for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

The developer noted the following:

- Risk-adjustment variables (15) include: age, heart rate >=150, SBP <=90, chronic renal, acute renal, GIB, cardiac arrhythmia, intracranial mass effect, mechanical ventilation, received CPR, cancer, cerebrovascular incident, cirrhosis, coma, medical admission or status post non-elective surgery, zero factor status (no risk factors other than age), and full code status (no restrictions on therapies or interventions at the time of ICU admission).
- The risk-adjustment model is based on the Intensive Care Outcomes Model - Mortality (ICOMmort).
- SDS variables were not available and were not used for reliability testing.
- The developer reported results at the patient and hospital level. At the patient level, the model had an AUC of 0.81, which the developer interprets as discrimination with good calibration. At the hospital level, the correlation coefficient between this model and the APACHE model (which was the model requiring the most data collection and with 3 times the time required to collect all variables, but which had an AUC of 0.89) was 0.92
- The developer does not include SDS factors in the risk model.
    - The developer performed an analysis on all 9,518 eligible patients admitted to the ICUs of 35 hospitals; time period not specified. For each patient, the developer had clinical risk factors and the following SDS variables: race, ethnicity, insurance status (including the categories of Medicaid and uninsured), and AHRQ Socioeconomic Status Index Score.
    - The developer reports none of the SDS variables were associated with mortality, even after accounting for differences in resuscitation preferences (Erickson S, et al. The Effect of Race and Ethnicity on Outcomes Among Patients in the Intensive Care Unit: A Comprehensive Study Involving Socioeconomic Status and Resuscitation Preferences. *Critical Care Medicine*, 2011; 39(3):429-35).
    - The developer speculates the lack of association between SDS and outcomes is because the entire course of care occurs within the hospital, where the mechanisms by which disparities occur (such as lack of access to care, food deserts, etc. may have less impact.)

Questions for the Committee:

- *Is the risk adjustment method appropriate?*
- *Do you agree with the developer's decision, based on its analysis, to exclude SDS factors in the risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer built a Bayesian model using hospital performance scores plus a random effect for each hospital. Hospitals were rated as different from expected if the coefficients on their variables were statistically significantly different from one.
- 95% confidence intervals on the point estimate of the outcome rate were calculated for each hospital using a Bayesian version of a hierarchical logistic regression model (implemented using WinBUGS). Using those confidence intervals, each individual hospital's performance was then compared to three benchmarks (the 10th, 50th, and 90th percentiles of performance among participating CHART hospitals). Hospitals were assigned to 1 of the 5 categories.
- Most recent quarter of public reporting—not identified more specifically by the developer—show:
  - 58.3% of hospitals would have been labeled as having "Average" performance
  - 6.2% would have been labeled "Superior"
  - 6.2% would have been labeled "Above Average"
  - 26.1% would have been labeled "Below Average"
  - 0.5% would have been labeled "Below Average".
- The developer states: "Hospitals believed the data were accurate and that low ratings were justified. The opportunity to participate in quality improvement collaboratives accompanied the reporting and were [sic] widely used by the hospitals. With this active response to the data, we saw a fairly dramatic decline in statewide ICU mortality. On subsequent audits, we confirmed that this was not due to increased coding of risk factors or any other manner of gaming."

Question for the Committee:

- *Does this methodology (observed to expected) identify meaningful differences **across** measured entities?*
- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- No data were missing, per the developer. The developer stated the web-based collection system includes data quality control that ensure all data are included and in range.

Guidance from the Validity Algorithm: 1→ 2 → 3 → 6→ 10 → 11 → 12 (eligible for MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Specifications are consistent with evidence.
**Correlation between trained auditors and hospital data collectors = 0.792.
% agreement between auditors and hospital data collectors = 94% (85-97%).
**The measure is consistent with the evidence

2a2. Reliability Testing

Comments:

**New validity testing at the data element-level: developer assessed agreement between trained auditors (the authoritative source) and hospital data collectors for all individual risk model elements
% agreement was 94% (RANGE 85-79%)
Questions for the Committee:
o The developer previously provided score-level testing and now provides data-element level testing. Do the results demonstrate sufficient validity so that conclusions about quality can be made? YES
**Sufficient validity demonstrated
**Same

**The validity testing results demonstrate sufficient reliability so that differences in performance can be identified. One caveat noted by the developers is the importance of quality data collection to ensure accurate, high-quality data are reported.
**This measure allows standardization in the formation of conclusions regarding quality

2b2. Validity Testing
Comments:
**2b3: Exclusions
Questions for the Committee:
o Are any patients or patient groups inappropriately excluded from the measure? NO
2b4: Risk adjustments--uses a statistical model
Questions for the Committee:
o Is the risk adjustment method appropriate? YES
o Do you agree with the developer's decision, based on its analysis, to exclude SDS factors in the risk-adjustment model? YES
2b5: Meaningful difference--uses a Bayesian model
Question for the Committee:
o Does this methodology (observed to expected) identify meaningful differences across measured entities? YES
o Does this measure identify meaningful differences in quality? YES
2b6: Comparability of data sources--not applicable
2b7: Missing data--not applicable
Guidance from the Validity Algorithm: 1→ 2 → 3 → 6→ 10 → 11 → 12 (eligible for MODERATE)
**Valid adjustment for SDS and case mix would be desirable. Unclear how hospital transfers are being handled.
**Exclusion: No patient groups appear to be inappropriately excluded from the measure evaluation.
Risk adjustment: appropriate
Meaningful Differences: This measure appears to be able to identify meaningful differences in quality performance.
**Exclusions are appropriate as they represent populations usually present in specialized centers or, specific databases.
Risk adjustment with an 0.81 AUC and 0.92 correlation with APACHE IV.
It provides meaningful differences in quality amongst evaluated entities

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Testing at both measure score and data element level
Guidance from the Reliability Algorithm: 1→ 2 → 3 → 6→7 → 8 (eligible for HIGH rating)
Question for the Committee:
o Do the results demonstrate sufficient reliability so that differences in performance can be identified? YES
**If I had a concern it would be the California centric aspect of the data utilized as it is unclear if this is nationally representative
**The reliability testing results demonstrate sufficient reliability so that differences in performance can be identified. One caveat noted by the developers is the importance of quality data collection to ensure accurate, high-quality data are reported.
**Differences in performance can be identified

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The measure requires manual chart abstraction because not all of the data elements are in defined fields in electronic sources. The developer is currently developing ICU eMeasures based on the Health Quality Measure Format (HQMF), to be completed in 2016. It anticipates 2 versions of this mortality eMeasure. The eMeasure also will incorporate ICU risk adjustment methodology.

- The developer reports each of 187 hospitals in California (full range of hospital type) submitted data on 400 patients/year (the first 100 consecutive patients in each quarter), and chart abstractors took ~11-15 minutes to collect data from each chart, which it considers minimal burden. The developer indicates an alternative sample size could be chosen by users, but it does not recommend a minimum sample size nor provide guidance on statistical power in this regard.
- There are no costs or licensing requirements.

Question for the Committee:

- *Does the usefulness of the measure outweigh the data collection burden of manual chart abstraction?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments:

**Requires chart abstraction but not reported to be burdensome. Developing ICU eMeasures.
Question for the Committee:
o Does the usefulness of the measure outweigh the data collection burden of manual chart abstraction? YES
**Data are being obtained from paper records.
**Agree this is feasible
**The data collection associated with these measures appears to be feasible. It will be more feasible when the eMeasures is implemented and will thus reduce the manual data extraction that is currently required. The usefulness of the measure outweighs the data collection burden of manual chart abstraction.
**This measure was developed for paper charts. an e-abstraction should be implemented now with "meaningful use" within the EHR.
This will allow a decreased burden. However, the current burden IS justified as it gives a reliable comparator for quality outcomes based on acuity.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

OR

Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details

- The developer notes that until 2013, this measure was used and publicly reported in California, but use was discontinued in favor of measures required by CMS.
- In 2013, developer began transforming this measure into an eMeasure for CMS consideration and anticipate this work to be completed in 2016. Currently, a model using data from two hospital EMRs is in progress.

Improvement results

The developer reports the following:

- A decline in ICU mortality from 13.85% to 11.67%, representing a change of -2.18%, was observed from 2007 to 2011 in a California sample that encompassed 69,483 patients and 187 hospitals. The data were collected as a part of the California Hospital Assessment and Reporting Taskforce (CHART) project.

Unexpected findings (positive or negative) during implementation

- The developer reports no challenges/difficulties with implementation, and no unexpected findings during implementation.

Potential harms

- The developer noted one potential unintended consequence: hospitals may seek to avoid high-risk patients. The developer noted such behavior could be monitored by evaluating changes in hospitals' risk profiles. The developer further stated that in the 4 years the measure was used for public reporting in California, this behavior was not observed or reported by hospitals

Feedback :

No feedback provided on QPS. The measure was not reviewed by MAP.

***Questions for the Committee**:*

- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *Since the developer is moving to an eMeasure and the measure is no longer being publicly reported, should the Committee be concerned about whether the measure in this format will be maintained going forward?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Measure is not currently publicly reported nor used in accountability programs. Plans to include in future accountability program.

Questions for the Committee:

o Do the benefits of the measure outweigh any potential unintended consequences? UNCERTAIN RE: HIGH RISK PATIENTS

o Since the developer is moving to an eMeasure and the measure is no longer being publicly reported, should the Committee be concerned about whether the measure in this format will be maintained going forward? UNTIL THE eMEASURE IS IN PLACE, THIS MEASURE IS APPROPRIATE.

POTENTIAL HARMONIZATION ISSUES WITH OTHER MEASURES

**I believe this is an important measure. However, with advancing trends in quality transparency, this metric carries the potential to lead to referral centers in urban areas that have a higher disease acuity being inaccurately represented compared with other institutions that will report data for this measure.

**The benefits of the measure outweigh any potential unintended consequences. The developer reports that this measure has been used extensively for accountability. The developers are moving to an eMeasure for CMS consideration. It seems appropriate to consider this measure for endorsement at this time. When the eMeasure (projected to be available later in 2016 for use in California) becomes available, perhaps its value can be assessed in the next round of review for this measure.

**Was publicly reported for a period of several years in California. The e-version should allow its inclusion in CMS. Could be added to Safety Scores like Leapfrog

Criterion 5: Related and Competing Measures

Related or competing measures

- 0702: Intensive Care Unit (ICU) Length-of-Stay (LOS)

Harmonization

- The measure specifications for #0702 and #0703 have been harmonized. The measures have been paired.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0703

Measure Title: Intensive Care: In-hospital mortality rate

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.

- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: **Intensive Care: In-hospital mortality rate**

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Prevention of death is the reason why patients are admitted to the ICU. The critically ill are a uniquely vulnerable patient population, and given the high costs of their care, determining which hospitals have higher than expected rates of mortality following ICU admission is a meaningful measure of ICU performance.

The observed mortality variation among ICU patients from different hospitals has prompted in-depth study of possible structural or process contributors to this variation. The field of health services research has sought to identify any number of predictors, which include, but are not limited to, presence of an ICU medical director, presence of an intensivist, nurse-to-patient ratio, use of ventilator weaning protocols, and even ICU teamwork factors. For instance, hospitals associated with an ACGME residency have been shown to perform better than expected when comparing observed to expected rates of death. Similarly interdisciplinary clinical rounds and the presence of an on-site emergency department have also been shown to be statistically significant variables in mortality prediction. These are but a few of the factors that have been pursued in an effort to identify what makes certain hospitals perform better than others when mortality is the measured outcome.

There is variable evidence about the link between structure and outcome, although in general it is believed that, for ICUs that are not in teaching hospitals, having more intensivist coverage is associated with better outcomes.

However, there are many evidence-based processes of care that do improve outcomes, including ventilator bundles, sepsis management, and appropriate choices of drugs (e.g., beta blockers or the correct antibiotic) or interventions (e.g., PCI for MI) for heart disease, infections, etc.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

In the process of evaluating possible structural or procedural factors contributing to mortality variation, a number of studies have been able to identify certain variables that are less significant. For instance, hospital size (>300 vs <300 patients) was not found to impact standardized mortality ratio in one particular study of 35 California hospital ICUs. Another study similarly looking at patient volume found that higher ICU patient volumes were associated with lower mortality rates, but only in high-risk critically ill adults. In a worldwide sample using the SAPS 3 database, authors found that factors related to nursing or physician staffing had no impact on performance ratings. Such findings may be specific to the populations to which the variables were studied, but warrant further examination.

In addition, there are many condition-specific processes—such as administering beta blockers or performing percutaneous coronary intervention for patients with acute myocardial infarction, or adhering to the ventilator bundle for patients that require mechanical ventilation—that are associated with improving the outcomes of ICU patients.

Glance LG, Yue L et al. Impact of patient volume on the mortality rate of adult intensive care unit patients. Crit Care Med 2006;34(7):1925-34.

Kuzniewicz MW, Vasilevskis EE, Lane R et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008 Jun;133(6):1319-27.

Rothen HU, Stricker K, Einfalt E et al. Variability in outcome and resource use in intensive care units. Intensive Care Med 2007;33:1329-36.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report
Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form 0703_ICU_Mort_evidence_attachment_2015.docx
1b. Performance Gap Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating: • considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or • disparities in care across population groups. 1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)* Prevention of death is the reason why patients are admitted to the ICU. The critically ill are a uniquely vulnerable patient population, and given the high costs of their care, determining which hospitals have higher than expected rates of mortality following ICU admission is a meaningful measure of ICU performance.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

A number of studies based on various large ICU patient samples have confirmed the marked variability in mortality outcomes. One of the earliest of such studies published in 1993 documented unadjusted in-hospital mortality rates from 42 intensive care units ranging from 6.4% to 40%, with 90% of this variation attributable to patient characteristics at admission. Corresponding observed to predicted mortality ratios varied from 0.67 to 1.25. Several studies have since observed similar wide variability among disparate samples. For instance, in a population of veterans, the mortality range was 2-30%, with observed to expected ratios ranging from 0.62-1.27. Again, in worldwide samples (as in the SAPS 3 database) and in geographically localized samples (California), mortality variability for patients admitted to the ICU persists.

Specifically for this measure, the most recent performance scores are based on data collected between the fourth quarter of 2010 and the third quarter of 2011. 226 hospitals contributed data, representing 69,483 patients and 8,112 deaths, for a overall unadjusted mean mortality rate of 11.67%. The standard deviation in this rate across hospitals was 4.17%, with an interquartile range of 8.56% to 13.68%; the minimum rate observed was 1.31%, while the maximum was 28.16%. The scores by decile are: 1.31% (0th percentile), 6.74%, 8.06%, 9.2%, 10.29%, 11.4%, 12.59%, 13.19%, 14.31%, 16.12%, and 28.16% (100th percentile).

These scores are contrasted with a baseline sample, which was collected from the first quarter to third quarter of 2007. For this baseline sample, the observed unadjusted mortality rate was 13.85%. Compared to the most recent sample, this represents a decline in mortality of 2.18% to 11.67%.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**

Knaus WA, Wagner DP, et al. Variations in mortality and length of stay in intensive care units. Ann Int Med 1993;118:753-61.
Kuzniewicz MW, Vasilevskis EE, Lane R et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008 Jun;133(6):1319-27.
Render ML, Kim M, Deddens J et al. Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure. Crit Care Med 2005;33(5): 930-9.
Rothen HU, Stricker K, Einfalt E et al. Variability in outcome and resource use in intensive care units. Intensive Care Med 2007;33:1329-36.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Among various populations, there are documented disparities in mortality outcomes following admission to the ICU. For instance, disease-specific racial variation has been noted among blacks, who have high mortality for critical care conditions such as sepsis or acute lung injury. In a similar fashion, the elderly have been looked at as a subpopulation in whom higher mortality rates occur after an ICU stay. In one study, older women fared worse than men in ICU outcomes. Age >65 together with mechanical ventilation >7d have together been demonstrated to be cofactors in predicting mortality. Aside from demographics, insurance status has also been implicated as a significant predictor, with the uninsured having higher risk of death.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

Barnato AE, Alexander SL et al. Racial variation in the incidence, care, and outcomes of severe sepsis: analysis of population, patient, and hospital characterisitics. Am J Respir Crit Care Med 2008 Feb 1;177(3):279-84.
Danis M, Linde-Zwirble WT et al. How does lack of insurance affect use of intensive care? A population-based study. Crit Care Med 2006 Aug;34(8):2235-6.
Erickson SE, Shlipak MH et al. Racial and ethnic disparities in mortality from acute lung injury. Crit Care Med 2009 Jan;37(1):1-6.
Feng Y, Amoateng-Adjepong Y et al. Age, duration of mechanical ventilation, and outcomes of patients who are critically ill. Chest 2009 Sept;136(3):157-64.
Fowler RA, Sabur N, Juurlink DN et al. Sex-and age-based differences in the delivery and outcomes of critical care. CMAJ 2007 Dec 4;177(12):1513-9.
Yang Y, Yang KS et al. The effect of comorbidity and age on hospital mortality and length of stay in patients with sepsis. J Crit Care

2009 Oct 14. [Epub ahead of print]

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Patient/societal consequences of poor quality, Affects large numbers, Severity of illness, Frequently performed procedure, A leading cause of morbidity/mortality, High resource use

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Between 1985-2000, the number of critical care beds in US acute care hospitals increased 26%. In addition, although the number of hospitals offering critical care has decreased, the proportion of total hospital beds assigned to critical care has increased. By 2005, critical care costs in the US were estimated to be $81.7 billion accounting for 13.4% of hospital costs, 4.1% of the national health expenditures and 0.66% of the gross domestic product. As a result, there is great interest in ensuring critical care is delivered at a high standard. One way in which the quality of ICU care can be measured is in-hospital mortality. As a quality indicator, mortality has been broadly accepted due to its relative ease of measurement and its importance to both clinicians and patients alike. Given the higher death rate of patients admitted to the ICU than those admitted to general hospital wards, mortality is an even more appropriate measure of outcome. In an effort to accurately predict risk of death, general ICU mortality risk-prediction models have been developed and refined over the last 20 years in order to better ´objectively´ describe ICU patient populations and use these descriptive variables to estimate patients´ risk for death. Patient characteristics, or case-mix, are significant contributors to risk of death, and risk-adjustment is prerequisite.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Gunning K, Rowan, K. ABC of intensive care: outcome data and scoring systems. BMJ 1999 Jul 24;317(7204):241-4.
Halpern NA. Can the costs of critical care be controlled? Curr Opin Crit Care 2009 Oct 9. [Epub ahead of print]
Halpern NA, Pastores SM et al. Changes in critical care beds and occupancy in the United states 1985-2000: differences attributable to hospital size. Crit Care Med 2006;34:2105-112.
Pronovost PJ, Miller MR et al. Developing and implementing measuers of quality of care in the intensive care unit. Curr Opin Crit Care 2001;7:297-303.
Rosenberg AL. Recent innovations in intensive care unit prediction models. Curr Opin Crit Care 2002;8:321-30.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://healthpolicy.ucsf.edu/sites/healthpolicy.ucsf.edu/files/documents/ICOM_Tool.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** ICU Outcomes Data Dictionary-633924321323431795.pdf

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
No changes have been made.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Total number of eligible patients whose hospital outcome is death. The measure is risk-adjusted, please see S.18.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Not-applicable; anyone with an ICU admission meeting eligibility criteria below is in the numerator.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Eligible patients include those with an ICU stay of at least 4 hours and >18 years of age whose primary reason for admission does not include trauma, burns, or immediately post-coronary artery bypass graft surgery (CABG), as these patient groups are known to require unique risk-adjustment. Only index (initial) ICU admissions are recorded given that patient characteristics of readmissions are known to differ.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Total number of eligible patients who are discharged (including deaths and transfers out to other hospitals).

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Eligible patients include those with an ICU stay of at least 4 hours and =18 years of age whose primary reason for admission does not include trauma, burns, or immediately post-coronary artery bypass graft surgery (CABG), as these patient groups are known to require unique risk-adjustment. Only index (initial) ICU admissions are recorded given that patient characteristics of readmissions are known to differ.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

<18 years of age at time of ICU admission, ICU readmission, <4 hours in ICU, primary admission due to trauma, burns, or immediately post-CABG, admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, transfers from another acute care hospital.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
<18 years of age at time of ICU admission (with time of ICU admission abstracted preferably from ICU vital signs flowsheet), ICU readmission (i.e. not the patient´s first ICU admission during the current hospitalization), <4 hours in ICU, primary admission due to trauma, burns, or immediately post-CABG, admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, patient transfers from another acute care hospital (i.e. patients whose physical site immediately prior to the index ICU admission was an acute care unit at an outside hospital)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not-applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Risk-adjustment variables include: age, heart rate >=150, SBP <=90, chronic renal, acute renal, GIB, cardiac arrhythmia, intracranial mass effect, mechanical ventilation, received CPR, cancer, cerebrovascular incident, cirrhosis, coma, medical admission or status post nonelective surgery, zero factor status (no risk factors other than age), and full code status (no restrictions on therapies or interventions at the time of ICU admission). The risk-adjustment model is based on the the Intensive Care Outcomes Model - Mortality (ICOMmort) with candidate interactions among variables and variable coefficients customized for the population of interest.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Model specifications are available online at:
http://healthpolicy.ucsf.edu/sites/healthpolicy.ucsf.edu/files/documents/ICU_Outcomes_Models_S9.pdf

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 hospital´s observed mortality rate and risk-adjusted mortality rate are both calculated using the abstracted data. For each hospital, the model produces a median and 95% confidence interval for the Standardized Mortality Ratio (SMR), which is the death rate for the hospital adjusted to the average case mix.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation*

Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
the first 100 consecutive eligible patients per quarter

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
In our implementation of this, missing data was not possible (because the hospital cannot close out and submit an incomplete file). If other implementation approaches were taken, for any missing variables we would recommend an assumption that the variable for which a value is missing is in the normal range (that is, the blood pressure was normal or the patient did NOT have a diagnosis risk factor such as cancer). This approaches incentivizes the hospital to provide complete and accurate data.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
ICU Outcomes Data Collection Instrument

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0703_ICU_Mortality_Testing_attachment_2015_1214_REV.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0703
Measure Title: Intensive Care: In-hospital mortality rate
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)

☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of

care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Disclaimer: the prior information included in this new form may not completely match with the layout of the new form.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data*

specified and intended for measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.23*)**	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

MPM III was developed on the Project IMPACT database of 124,855 patients treated in 135 ICUs at 98 hospitals between 10/2001 and 3/2004. The majority of the hospitals were in the United States. Although reliability of data collection was ensured by the developers of the MPM III model, analyses and statistics were not reported in their original publication (Crit Care Med 2007;35:827-35). After 200 records were excluded due to missing essential MPM or outcome variables, 124,885 patients were available for analysis and model development. 50,307 (40.3%) of the patients were used for model validation.

The subsequent reliability testing methods and results are taken from Chest 2008 Jun;133(6):1319-27, in which the data sample was comprised of 11,300 patients admitted to 35 California hospital ICUs from 2001-2004.

(answering new question 1.2.)

Original reliability testing methods and results are taken from *CHEST* 2008 Jun;133(6):1319-27, in which the data sample was comprised of 11,300 patients admitted to 35 California hospital ICUs from 2001-2004. The models were then updated each year in our work with the California Hospital Assessment and Reporting Taskforce. In this program data were collected on ~70,000 patients per year. The reliability testing described below reflects audits performed among hospitals participating in the CHART program.

1.3. What are the dates of the data used in testing? October 2010-September 2011

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)) **(this was not a separate question before)**

187 hospitals of all sizes in California were in the pool to potentially be selected for random audits (more details in 1.7 below). Hospitals included teaching and non-teaching institutions, rural (including Critical Access Hospitals) and urban institutions.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*) **(this was not a separate question before; see above answers in blue)**

Almost all patients ≥18yo admitted to ICUs who stayed at least 4 hours were included. Exclusion criteria were: 1) having a primary reason for admission of trauma, burns, and coronary artery bypass graft (CABG) surgery patients (because other performance measures exist for these populations), 2) admitted to exclude myocardial infarction (MI) and subsequently found without MI or any other acute process requiring ICU care, and 3) transfers from another acute care hospital. Only index (initial) ICU admissions are recorded given that patient characteristics of readmissions are known to differ. Collectively, the participating hospitals had 76% of all ICU patients in California. Each hospital was asked to submit data on 400 patients a year (the first 100 consecutive patients in each quarter).

Discrimination and calibration statistics for the risk adjustment model were calculated on the entire population. Auditing of the reliability of critical data elements involved having a second reviewer re-abstract all charts and was performed on 986 patients from (34 each from 29 hospitals selected randomly from each performance group with intentional over-sampling of hospitals rated as performing better or worse than expected in terms of ICU mortality, since this was the outcome expected most likely to induce gaming—if any gaming were to occur). This approach gave us 60% power to detect hospitals that were truly misclassified in terms of ICU mortality performance.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

1.8 **What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used?** For example, patient-reported data (e.g., income, education, language), proxy

variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate). **(new question)**

SDS variables were not available and were not used. See section 2b4.4b for evidence that race, ethnicity, insurance status and AHRQ Socioeconomic Status Index Score variables were not associated with ICU outcomes.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data were reabstracted by auditors on a 5% random sample of patients. Kappa statistics were calculated for inter-rater variability between the data abstractor and the auditor. The auditors were clinical nurses who were trained by the authors and completed extensive sample chart abstraction.

See section 2b2 for validity testing of individual data elements.

For the reliability testing of the performance score, we calculated the correlation between probability of death calculated on the data collected by the hospital's data collector and the probability of death calculated on the data collected by trained auditors (blinded to the data originally submitted by the hospital's data collector). This provides a sense of how closely the scores agreed, but does not provide (for any disagreement), any indication of whether the hospital was reporting higher or lower severity than the auditor found.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

For physiologic variables of the MPM III mortality prediction model, inter-rater reliability was excellent, with agreement ranging from 91.5 to 98.8%, and weighted kappa statistics ranging from 0.72-0.96.

See section 2b2 for validity testing of individual data elements.

The correlation coefficient between the hospital's predicted probabilities of death and the auditor's predicted probabilities was 0.792. There was no clear pattern suggesting that the hospitals over-reported risk factors (that is, in some cases, hospitals were over-reporting, in others, they were under-reporting).

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The data collection for this measure requires training and mistakes can be made both in favor of and against the reporting hospitals. Periodic audits and training of data collection staff are recommended. On the whole, though, the errors did not seem to reflect gaming of the system (though this was only tested in a context of

public reporting, not a payment). Thus, we conclude that, like all other measures, it is likely that attention needs to be paid to data quality.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

ORIGINAL ANSWER MOVED TO QUESTION 1.2 OF THIS FORM

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Model performance in the validation sample was tested using area under the receiver operating characteristics curve (ROC) >=0.75 as a measure of discrimination and the Hosmer-Lemeshow goodness-of-fit statistic as a measure of calibration. Acceptable calibration was defined as: a) a nonsignificant Hosmer-Lemeshow value; b) a Hosmer-Lemeshow decile calibration plot with a slope and intercept not differing significantly from 1 and 0, respectively; and c) the SMR on the validation set between 0.95-1.05 with its confidence intervals including 1.

For the validity tests conducted, we used the kappa statistic, which expresses how much better or worse the agreement between two observers was compared to chance alone. Testing was conducted on 986 patients from (34 each from 29 hospitals selected to over-sample hospitals rated as performing better or worse than expected, with 60% power to detect hospitals that were truly misclassified).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Area under the ROC curve was 0.823 (95% CI 0.818-0.828), the Hosmer-Lemeshow statistic 11.62 (p = 0.31), and the standardized mortality ratio was 1.018 (95% CI 0.996-1.040). Actual mortalities closely tracked MPM III predictions by deciles of predicted risk.

Percent agreement between auditors and hospital data collectors across all individual risk model elements was 94%, with a range for specific risk variables from 85-97%. Kappas for specific variables ranged from 0.37-0.86, with most above 0.6 (substantial agreement or better). The only variable below Koch and Landis' criteria for at least moderate agreement was the coma variable (kappa 0.37), with the auditors identifying more cases than the hospitals had coded.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*) **(new question)**

The data collection for this measure requires training and mistakes can be made both in favor of and against the reporting hospitals. Periodic audits and training of data collection staff are recommended. On the whole, though, the errors did not seem to reflect gaming of the system (though this was only tested in a context of public reporting, not a payment). Thus, we conclude that, like all other measures, it is likely that attention needs to be paid to data quality.

{This is Landis and Koch's interpretation scale for kappas: "The kappa statistic has a maximum value of 1 when agreement is perfect, zero when agreement is no better than chance. The nomenclature of Landis and Koch is usually used to interpret the degree of agreement: kappa <0.21, slight agreement; 0.21-0.40, fair agreement; 0.41- 0.60, moderate agreement; 0.61-0.80, substantial agreement; and 0.81-1.00, almost perfect agreement." [Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174]}

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

Trauma, burn, and coronary artery bypass surgery patients (whose outcomes are assessed using other methods—e.g., the Society of Thoracic Surgeons' bypass surgery mortality reports), were excluded.

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not-applicable

Because these exclusions were pre-specified before the data collection began, no testing of the exclusions has been performed.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not-applicable

Not applicable.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not-applicable

Not applicable.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 15 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

ORIGINAL ANSWER TO ORIGINAL 2b4.1 QUESTION MOVED TO QUESTION 1.2 OF THIS FORM

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed

to achieve fair comparisons across measured entities.

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

Data analyzed were from ICUs with >= 100 patient records in the Project IMPACT database. The MPM II variables were used in the MPM III update, and new candidate variables were also evaluated. Univariate analysis assessed the relationship of the MPM II independent variables to mortality using Student's t-tests and chi-squared tests with a significance level of 0.05. Multivariate logistic regression with robust variance estimators was performed using variables with a significant univariate relationship to outcome. Interactions were considered (particularly between age and other MPM variables) since initial calibration suggested that age effects were influenced by the presence of comorbidities.

2b4.3. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **(new question)**

In building this model, we first considered all variables in all previously published ICU mortality risk adjustment models (see *CHEST* 2008 Jun;133(6):1319-27). With stakeholder input, we compared the models' predictive accuracy (discrimination and calibration) vs the cost of data collection (using chart abstractor time and motion studies). As it turned out, hospital performance ratings varied little among the different risk adjustment models, but the cost of data collection varied substantially. Therefore, the stakeholders chose the model that minimized data collection cost.

2b4.4a. What were the statistical results of the analyses used to select risk factors? (re-worded question on results)

Not-applicable

At the patient level, the model had an AUC of 0.81 for discrimination with good calibration. At the hospital level, the correlation coefficient between this model and the APACHE model (which was the model requiring the most data collection and with 3 times the time required to collect all variables, but which had an AUC of 0.89) was 0.92.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects) (new question)

We do not include SDS factors in our ICU models. In the development work for these models, we performed an analysis on all 9,518 eligible patients admitted to the ICUs of 35 hospitals (using the same eligibility criteria as are proposed for this measure). For each patient, we had, in addition to the clinical risk factors, the following SDS variables: race, ethnicity, insurance status (including the caterogies of Medicaid and uninsured), and AHRQ Socioeconomic Status Index Score. *None of the SDS variables were associated with mortality* (Erickson, S, Vasilevskis, EE, et al. The Effect of Race and Ethnicity on Outcomes Among Patients in the Intensive Care Unit: A Comprehensive Study Involving Socioeconomic Status and Resuscitation Preferences. *Critical Care Medicine*, 2011; 39(3):429-35). This was true even after accounting for differences in resuscitation preferences. We suspect that the lack of association between SDS and outcomes seen here is because the entire course of care occurs within the hospital, where the mechanisms by which disparities occur (such as lack of access to care, food deserts, etc) may have less impact.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*) **(new question)**

Discrimination was assessed using AUCs, calibration using both Hosmer-Lemeshow statistics and calibration curves.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

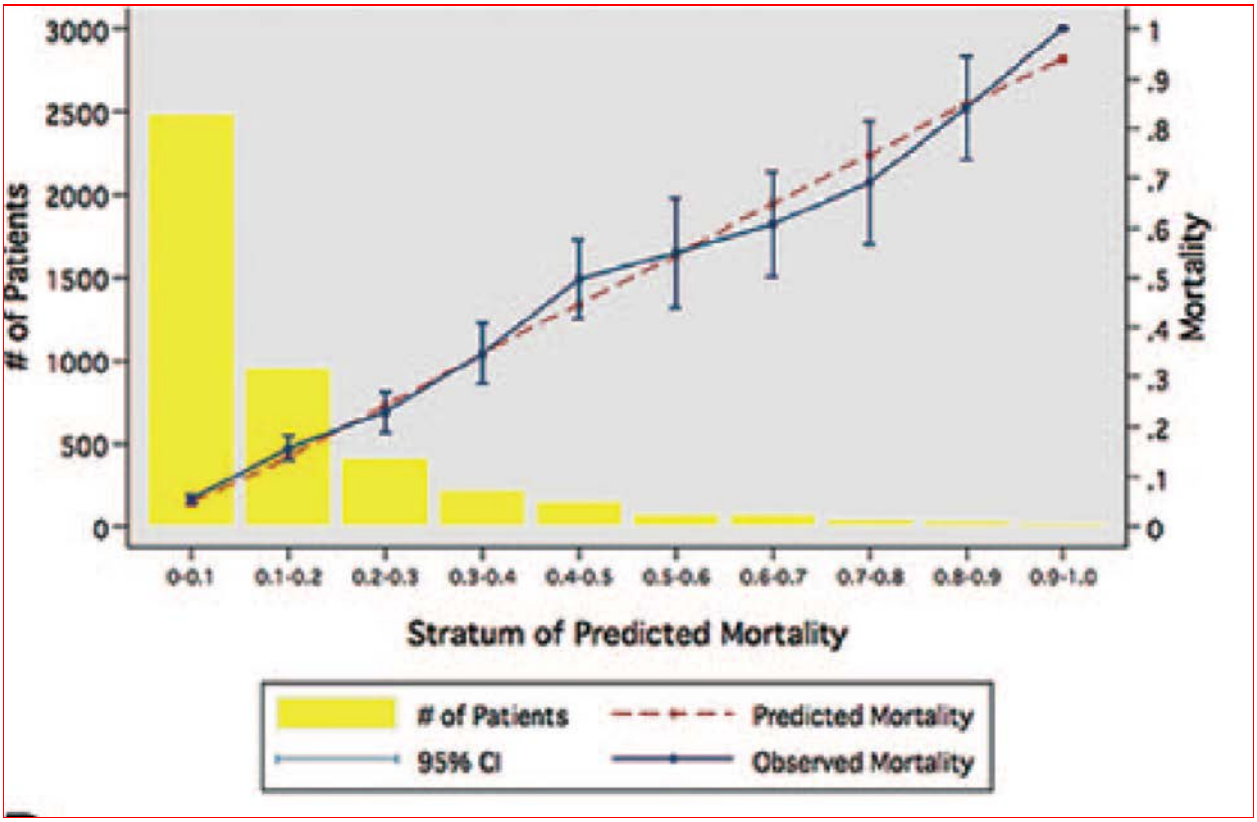
2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**: (new question)**

c-statistic (AUC) of 0.81 for discrimination

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): **(new question)**

H-L statistics 9.8 (not significant)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
(new question)



2b4.9. Results of Risk Stratification Analysis: **(new question)**

Not applicable.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*) **(new question)**

The model demonstrates adequate discrimination and calibration.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*) **(new question)**

All hospitals agreed to be reported using this approach. Despite extensive media coverage of the resulting public report, no hospital has ever complained that the data or the model were inaccurate.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

A Bayesian model was built using hospital performance scores plus a random effect for each hospital. Hospitals were rated as different from expected if the coefficients on their variables were statistically significantly different from one.

In order to evaluate the performance of the customized MPM III model in our population of interest, the AUROC was calculated and found to be 0.829 (a value comparable to that found by model developers.) Moreover, when hospital SMR rankings using MPM III were compared to those generated using MPM II, the Spearman rank correlation coefficient was 0.97, indicating a strong correlation in hospital rankings (i.e. performance measurement).

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Using the customized version of the MPM III on the current database of California hospitals yielded the following summary statistics of the hospital SMRs:
Mean 0.989293
Median 0.988061
Std Deviation 0.26515
Interquartile Range 0.83-1.13

(answering new question)
Our public reporting program, CHART uses five levels of ratings: “Superior”, “Above Average”, “Average”, “Below Average”, and “Poor”. As a first step, 95% confidence intervals on the point estimate of the outcome rate were calculated for each hospital using a Bayesian version of a hierarchical logistic regression model (implemented using WinBUGS). Using those confidence intervals, each individual hospital's performance was then compared to three benchmarks (the 10th, 50th, and 90th percentiles of performance among participating CHART hospitals). Hospitals were assigned to 1 of the 5 categories using the table below. CHART Methodology Table.

1. CHART Group Calculation

a) Using statistical model with hospital rates as random effects, using the program WinBUGS, hospital effect point estimates and 95% confidence intervals were calculated.

b) A low, middle and high reference value was obtained, as the 10th, 50th, and 90th weighted percentile rate. By definition, 10% of all patients were in hospitals whose rate was less than or equal to the 10th weighted percentile. Similarly for the 50th and 90th weighted percentiles.

2. Mortality Performance Group calculation

A hospital's confidence interval was compared to the low, middle, and high reference values to determine the CHART Group as follows:

"L" represents the location of the hospital's lower confidence limit
"U" represents the location of the hospital's upper confidence limit

2. Mortality Performance Group	2. Mortality Performance Group	2. Mortality Performance Group	2. Mortality Performance Group	2. Mortality Performance Group
Below Low Reference	Between the Low and Middle	Between Middle and Top	Above Top Reference	CHART Performance Group
L & U				1
L	U			1
	L & U			2
L		U		2
	L	U		3
		L & U		3
	L		U	4
		L	U	4
			L & U	5

Mortality Reference Values		
Low	Middle	High
6.87%	11.47%	15.75%

3. CHART Performance Group Scoring Key

1 - Superior
2 - Above Average
3 - Average
4 - Below Average
5 - Poor

New ratings were generated every quarter. In the most recent quarter of public reporting, 58.3% of hospitals were labeled as having “Average” performance. Of the remaining hospitals, 6.2% were labeled “Superior”, 6.2% were labeled “Above Average”, 26.1% were labeled “Below Average”, and 0.5% were labeled “Poor”. This represented a skew toward more negative labels, but we believe this reflected that some hospitals were more actively engaged in improvement (so that smaller group was at the top while a larger group of hospitals that had not yet worked hard on ICU care was on the lower end of performance). Of note, despite this negative skew, very few hospitals expressed to us that the measurement system might be unfair and no hospitals complained to the media that the data were inaccurate or the performance measurement system unfair.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
(new question)

Hospitals believed the data were accurate and that low ratings were justified. The opportunity to participate in quality improvement collaboratives accompanied the reporting and were widely used by the hospitals. With this active response to the data, we saw a fairly dramatic decline in statewide ICU mortality. On subsequent audits, we confirmed that this was not due to increased coding of risk factors or any other manner of gaming.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not-applicable

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not-applicable

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
(new question)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
(new question)

All data were submitted through a secure website that performs data quality control during the submission process, so no data were missing (each chart cannot be closed with missing or out of range data).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
(new question)

Not applicable.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
(new question)

Not applicable.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a**

credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.
We are currently in the process of developing ICU eMeasures based on the Health Quality Measure Format (HQMF). We anticipate that there will be two versions of this mortality eMeasure: one to compute the raw (unadjusted) mortality rate and another to extract the data elements to compute risk-adjusted rates. The eMeasure will also incorporate ICU risk adjustment methodology that is currently under review for publication. We anticipate that development of this eMeasure will be completed in 2016.

The HQMF for this eMeasure will be developed using the Measure Authoring Tool based on the Quality Data Model (QDM). In addition, we anticipate that we will also be able to define Category I-III Quality Reporting Document Architecture (QRDA) documents with the eMeasure HQMF.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
In 187 hospitals in California (from small rural hospitals to the largest teaching hospitals), we have successfully collected this data. The average time per chart for an experienced data collector was 11-15 minutes. We collected data on 100 patients per quarter to minimize the data collection burden while still getting sufficient sample size to get precise estimates of hospital performance. However, an alternative target sample size could easily be chosen by users.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
No fees or licensing are involved.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	

Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

These measures were used and publicly reported in California until 2013. At that time, the decision was made to focus on measures required by CMS. Therefore, we are currently working to transform this model into an e-measure for CMS consideration.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

This has already been used extensively for accountability. However, we intend to transform to e-measures and have made substantial progress in that direction. We have been able to implement a model that uses data from 2 hospital EMRs and has superior predictive performance to the existing model. A paper describing this is under review. However, we have not yet transformed this model into HQMF. We expect this to be completed in 2016.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

A decline in ICU mortality from 13.85% to 11.67%, representing a change of -2.18%, was observed from 2007 to 2011 in a California sample (see 1b.2, “performance scores over time”). 69,483 patients and 187 hospitals were represented in these data, which was collected as a part of the California Hospital Assessment and Reporting Taskforce (CHART) project.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

The potential unintended consequence is that hospitals may seek to avoid high-risk patients. One could monitor this behavior by evaluating changes in hospitals´ risk-profiles over time. However, in the four years that this model was used for public reporting in California, no such behavior was observed or reported by hospitals (and the program had a very active clinical group).

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0702 : Intensive Care Unit (ICU) Length-of-Stay (LOS)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses 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.**
This measure is completely harmonized with measure 0702: Intensive Care Unit (ICU) Length-of-Stay (LOS)

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Philip R. Lee Institute for Health Policy Studies
Co.2 **Point of Contact:** R. Adams, Dudley , adams.dudley@ucsf.edu , 415-476-8617-
Co.3 **Measure Developer if different from Measure Steward:** Philip R. Lee Institute for Health Policy Studies
Co.4 **Point of Contact:** R. Adams, Dudley , adams.dudley@ucsf.edu , 415-476-8617-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Not-applicable

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 1990
Ad.3 **Month and Year of most recent revision:** 10, 2011
Ad.4 **What is your frequency for review/update of this measure?** annually when in use
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.6 **Copyright statement:** is in the public domain
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0708

De.2. **Measure Title:** Proportion of Patients with Pneumonia that have a Potentially Avoidable Complication (during the episode time window)

Co.1.1. **Measure Steward:** Health Care Incentives Improvement Institute

De.3. **Brief Description of Measure:** Brief Description of Measure: Percent of adult population aged 18+ years with Community Acquired Pneumonia who are followed for one-month, and have one or more potentially avoidable complication (PAC) during the episode time window. Please reference the attached document labeled NQF_PNE_all_codes_risk_adjustment_12_14_15.xls, in the tab labeled PACS I-9 & I-10 for a list of code definitions of PACs relevant to pneumonia.

Community Acquired Pneumonia may be managed in an inpatient setting, where the patient is admitted to a hospital within 1-3 days of onset of symptoms, or in milder cases, patients may be hospitalized a little later in the course of illness, or never at all where management could be solely in an outpatient setting. In any of these circumstances, potentially avoidable complications (PACs) may occur during the index stay, in the post-discharge period; or in patients who were never hospitalized, PACs may occur any time during the episode time window. Readmissions due to pneumonia or due to any related diagnosis are also considered as PACs.

We define PACs as one of two types:
(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a type 1 PAC if they develop one or more complication directly related to pneumonia or its management. Examples of these PACs are respiratory insufficiency, other lung complications, fluid electrolyte acid base problems, sepsis, respiratory failure etc.
(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are considered to have a type 2 PAC, if they develop any of the complications related to patient safety failures such as phlebitis, deep vein thrombosis, pressure sores or for any of the CMS-defined hospital acquired conditions (HACs).

PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PAC in any of the above settings, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_PNE_all_codes_risk_adjustment_12_14_15.xls serves as an example. The tab labeled PAC overview gives the percent of pneumonia episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in pneumonia episodes within this dataset.

The information is based on a two-year claims database from a large regional commercial insurer. The database had 3,258,706 covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

1b.1. **Developer Rationale:** Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of potentially avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications

as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, hospital, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 60% of its plan members with pneumonia incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow-up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events" and non-payment for certain readmissions) and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x
2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.
3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.
4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt
5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.
6) Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf
7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.
8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html
9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.
10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/
11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.
12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.
13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional

Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.
14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.
15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.
16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.
17) Leibson CL1, et al. "Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort." Med Care 46.2 (2008):127-32.

S.4. Numerator Statement: Outcome: Number of patients with pneumonia who had one or more potentially avoidable complications (PACs) during the episode time window.
S.7. Denominator Statement: Adult patients aged 18 years and above who have a pneumonia episode and are followed for at least one-month.
S.10. Denominator Exclusions: The target population captures adult patients (18+) in the dataset, who have a complete episode of community-acquired pneumonia, with no enrollment gaps, and no outlier costs. Patients who do not meet these criteria are excluded from the target population.

De.1. Measure Type: Outcome
S.23. Data Source: Administrative claims
S.26. Level of Analysis: Clinician : Individual, Facility, Population : Regional

IF Endorsement Maintenance – Original Endorsement Date: Jan 17, 2011 **Most Recent Endorsement Date:** Jan 17, 2011

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- Developer attests there is new evidence since the last submission.
- The developer provides the following information for this outcome measure:
  - Pneumonia is a leading cause of mortality and morbidity in the U.S. population (CDC 2015). There may be up to 5 to 6 million cases of community-acquired pneumonia (CAP) diagnosed annually in the United States, accounting for approximately 1 million hospitalizations and approximately 10 million physician visits (Aliberti 2008).
  - The updated evidence includes studies that identify ways to reduce PACs at the healthcare provider level for an individual with pneumonia:
    - Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
    - DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)
    - Frequent change in position of AMI patients in the CCU avoids pressure sores (Sullivan 2013)
    - Better processes of care significantly reduce occurrence of potentially avoidable complications (PACs) in all settings – while the patient is in the hospital and post-discharge (Wachter 2013).

Questions for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale appears to be the same since the last NQF endorsement review. The evidence provided by the developer on ways to reduce PACs to indicate is updated and directionally the same. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following:

- Data encompassed 2 two years of administrative claims for the period April 1, 2012 through December 17, 2014. The data set included 13,228 episodes of pneumonia, with more than 54% of episodes having one or more potentially avoidable complications.
- The developer reports performance scores (PAC rates) for the 82 facilities and 170 providers that had at least 10 patients:

Facility Unadjusted PAC Rates (n=82)
Median (IQR): 63% (58%, 69%)
Range: 27%-100%

Facility Risk-Standardized PAC Rates (RSPR)
Median (IQR): 63% (57%, 69%)
Range: 30%- 1%

Physician Unadjusted PAC Rates (n=170)
Median (IQR): 60% (43%, 79%)
Range: 0%-100%

Physician Risk-Standardized PAC Rates (RSPR):
Median (IQR): 58% (44%, 70%)
Range: 0%-100%

Disparities

- The developer did not submit disparities data. It did, however, summarize data from literature that address disparities:
  - Some studies demonstrate that African American patients with community acquired pneumonia (CAP) do not receive timely antibiotic treatment or lag behind in other process measures such as smoking cessation counseling or pneumococcal or influenza immunizations (Fine 2002, Mortensen 2004, Bennett 1995, Hausmann 2009).
  - Conversely, a large multicenter study based on an analysis of more than 40,000 patients with CAP managed over 5 years across 150 VA hospitals found African Americans and Caucasians were equally likely to receive guideline-concordant antibiotics and experienced similar 30-day mortality when treated in medical wards. Additionally, when admitted to the ICU, African Americans experienced a survival advantage with a lower 30-day mortality as well as shorter hospital LOS as compared to their Caucasians counterparts (Frei 2010).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of additional evidence related to disparities in this area of healthcare?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The developer provides the following information for this outcome measure:

Pneumonia is a leading cause of mortality and morbidity in the U.S. population (CDC 2015). There may be up to 5 to 6 million cases of community-acquired pneumonia (CAP) diagnosed annually in the United States, accounting for approximately 1 million hospitalizations and approximately 10 million physician visits (Aliberti 2008).

The updated evidence includes studies that identify ways to reduce PACs at the healthcare provider level for an individual with pneumonia:

Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)

DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)

Frequent change in position of AMI patients in the CCU avoids pressure sores (Sullivan 2013)

Better processes of care significantly reduce occurrence of potentially avoidable complications (PACs) in all settings – while the patient is in the hospital and post-discharge (Wachter 2013).

**This is an outcomes measure and the evidence relates to the outcome being measured.

**is there at least on thing provider can do to acheive a change? YES

does committee agree no need fro repeat discussion on evidence? YES

**Given how broad the outcome measure is, there is no question that there is at least one thing a provider could do to change the measure result.

Agree that a lack of change would support the idea that we do not need to revote. Should there be a review of whether the list of PACs should be re-examined?

**Strong evidence that supports relationship between interventions/actions/inactions and outcomes measured by the implementation of the quality measure

** Outcome measure – occurring within 30d of "CAP"

Type 1 PAC – related to pneumonia

Type 2 – HAC

Numerator = patients with pneumonia with one or more PAC –

Denominator = age at least 18 with pneumonia episode followed for at least one month (or is it 30 days)?

How does this measure add to measures regarding readmissions and HACs?

There is a lot of overlap with existing measures and I am unclear what this one adds. There is also language that raises concerns that it is not completely harmonized with HAC specifications.

1b. Performance Gap

Comments:

**Data encompassed 2 two years of administrative claims for the period April 1, 2012 through December 17, 2014. The data set included 13,228 episodes of pneumonia, with more than 54% of episodes having one or more potentially avoidable complications. The developer reports performance scores (PAC rates) for the 82 facilities and 170 providers that had at least 10 patients:

Facility Unadjusted PAC Rates (n=82)
Median (IQR): 63% (58%, 69%)
Range: 27%-100%

Facility Risk-Standardized PAC Rates (RSPR)
Median (IQR): 63% (57%, 69%)
Range: 30%- 1%

Physician Unadjusted PAC Rates (n=170)
Median (IQR): 60% (43%, 79%)
Range: 0%-100%

Physician Risk-Standardized PAC Rates (RSPR):
Median (IQR): 58% (44%, 70%)
Range: 0%-100%

Disparities

The developer did not submit disparities data. It did, however, summarize data from literature that address disparities:

Some studies demonstrate that African American patients with community acquired pneumonia (CAP) do not receive timely antibiotic treatment or lag behind in other process measures such as smoking cessation counseling or pneumococcal or influenza immunizations (Fine 2002, Mortensen 2004, Bennett 1995, Hausmann 2009).

Conversely, a large multicenter study based on an analysis of more than 40,000 patients with CAP managed over 5 years

across 150 VA hospitals found African Americans and Caucasians were equally likely to receive guideline-concordant antibiotics and experienced similar 30-day mortality when treated in medical wards. Additionally, when admitted to the ICU, African Americans experienced a survival advantage with a lower 30-day mortality as well as shorter hospital LOS as compared to their Caucasians counterparts (Frei 2010).
**There were no specific or consistent performance gaps noted.
**is there a gap in performance? YES
**There is a high rate of PACs. Not sure that such supports that there is a "gap" in care.
I am not aware of any other data concerning disparities.
**Data supports performance gaps; disparate studies summarized speak to pneumonia, but were there studies of PACs incidences and if there are disparate populations for that?
Also, in the testing completed, did the reliability testing not include parsing the data to see if any differences were identifiable in that data set?
** No disparities data were presented.

1c. High Priority (previously referred to as High Impact)
Comments: **NA
**Not applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

- The specifications have changed since the last submission:
  - Code tables have been revised to display codes as either I-9 or I-10 codes.
  - Only Community Acquired Pneumonia codes have been retained as potential trigger codes.
  - All codes have been updated to 2015 (current codes) and ICD-10 code conversions.
  - PAC definitions are based on diagnosis codes
  - PACs are not defined with procedure codes, but are now based on diagnosis codes
  - Instead of 3 types of PACS, there are 2: Type 1 PACs (related to index condition) and Type 2 PACs (safety failure)
  - Service assignment logic has been modified.
  - Databases have been expanded to include the Medicaid population.
- The numerator of this measure is: *Number of patients with pneumonia who had one or more potentially avoidable complications (PACs) during the episode time window*.
- The denominator of this measure is: *Adult patients aged 18 years and above who have a pneumonia episode and are followed for at least one month.*
- The ICD-9 and ICD-10 codes have been included in the specification details.
- The calculation algorithm is stated in S.18.
- This measure is risk adjusted.

Questions for the Committee :

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- The developer states it tested the reliability of the PAC calculations on 2 databases, but "no formal reliability testing done."
    - A database with 87,500 members with total claims volume equaling 2.3 million records; 70.5% of patients with Pneumonia incurred PACs
    - A regional health plan with more than 500,000 plan members; 71% of patients with Pneumonia incurred PACs
- The developer merely reports "these results were very consistent with the analysis in our national database."

Describe any updates to testing

- The developer indicates reliability testing has been updated.
- Specifically, testing of the measure score was conducted on a administrative claims dataset (April 1, 2012 to December 17, 2014) from a large regional commercial insurer. The data include medical and pharmacy claims on more than 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.
- Testing encompassed included 82 facilities, 170 physicians, and 13,228 episodes of pneumonia. Patients in these episodes were, on average, 57.1 years of age (range 18-64 years) and 45.7% were female. Race information was not included.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developer uses a beta-binomial method to demonstrate that the measure sufficiently differentiates performance among providers.
- The developer states there is not a clear cut-off for minimum reliability level, but citing Adams (2009), notes values above 0.7 are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians or pairs of facilities.

Results of reliability testing

- The developer reports the following reliability testing results for facilities:
    - Facilities with at least 10 episodes: 25% of scores above the minimum accepted level of 0.7.
    - Facilities with at least 80 patients: median reliability score was 0.7
    - Facilities that exceeded 125 attributed patients: reliability scores consistently above 0.70
    - The developer concluded that for facilities with a high number of episodes, the pneumonia PAC measure sufficiently differentiates facility performance
- The developer reports the following reliability testing results for physicians:
    - Physicians with at least 10 episodes: more than 75% reliability scores of 0.78 or higher; 10% of physicians in the sample had scores below 0.70.
    - The developer concluded the measure differentiates performance across a wide range of sample sizes.
- The tables below provide summaries and additional detail of the reliability analyses.

Facilities

Reliability Scores	Minimum # Episodes Per Facility	
	>=10	>=125
# of Providers (%)	82 (100)	15 (18)
Median (IQR)	0.56 (0.34,0.69)	0.78 (0.73,0.82)
Range	0.16-1.00	0.71-0.84

Physicians

Reliability Scores	Minimum # Episodes Per Provider
	>=10
# of Providers (%)	170 (100)
Median (IQR)	0.85 (0.78,0.92)
Range	0.66-1.00

- The table below compares the minimum sample sizes required for the various PAC measures to achieve an "absolute" and "median" reliability score of > 0.7 for both facilities and physicians.

Unit	Sample Size for "Absolute" Reliability >0.7	Sample Size for "Median" Reliability >0.7
Facility	125	80
Physician	11	10

- The developer reports that all missing cases are deleted from the numerator and denominator.
- The developer reports the following reliability testing results for facilities:
  - Facilities with at least 10 episodes: reliability scores were generally low, with only 25% of scores above the minimum accepted level of 0.7.
  - Facilities with at least 80 patients: median reliability score was 0.7
  - Facilities that exceeded 125 attributed patients: the reliability scores were consistently above 0.70
  - The developer concluded that for facilities with a high number of episodes the pneumonia PAC measure differentiates facility performance
- The developer reports the following reliability testing results for physicians:
  - Physicians with at least 10 episodes: more than three-quarters had reliability scores of 0.78 or higher, well above 0.70. Moreover, just 10% of physicians in the sample had scores below this threshold.
  - The developer concluded the measure differentiates performance across a wide range of sample sizes.

Guidance from the Reliability Algorithm: 1→ 2 → 4→ 5→ 6 (highest eligible rating is HIGH)

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified for facilities? for physicians?*
- *Should the Committee recommend the developer specify a minimum sample size directly in the specifications?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity. PAC analysis presented to physicians in 3 different healthcare systems, as well as medical directors from the employer coalition that provided the dataset to run the analyses. The developer stated the

face validity testing was positive, and the physicians to whom the analyses were shown concurred with the construction of the PACs and their definitions. The developer did not report specifically on a face validity assessment at the level of the measure score, as required by NQF.

Describe any updates to validity testing

- The developer presents significantly expanded face validity in the current submission (i.e., number of individuals and types of entities). Although a systematic assessment is not presented, the developer states it has not had results rejected in whole or part. Based on the reference to "results," NQF staff infer this indicates face validity at the performance score level.

SUMMARY OF TESTING

Validity testing level ☒ Measure score ☐ Data element testing against a gold standard ☐ Both

Method of validity testing of the measure score:

☒ **Face validity only**
☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer reports assessing face validity through application as follows:
    - Claims data analyses to physicians in different healthcare systems
    - Pennsylvania Employee Benefits Trust Fund, local provider groups and hospitals, Horizon Blue Cross Blue Shield of NJ, and physicians and health systems
    - Analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:
        - Vermont Payment Reform Commission
        - Maine Health Management Coalition
        - WellPoint / Anthem CT
        - NY State Medicaid
        - CT Medicaid
        - CO All-payer Claims Database, Center for Improving Value in Health Care

Validity testing results:

- The developer states it has not had results "rejected in whole or part," which NQF staff infers as face validity at the performance score level.

Question for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer reports the following denominator exclusions:
    - Patients < 8 years
    - Patients whose gender is missing
    - Patients who do not have continuous enrollment for the entire time window with the entity providing the data
    - Patients whose pneumonia episode time window extends outside the dataset time period
    - The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type)
    - "Claims" are excluded from the pneumonia measure if they are considered not relevant to pneumonia care
- The developer did not complete formal exclusion testing or formal analysis on the impact of exclusions on performance scores as requested by NQF.

Questions for the Committee:

- *Are the exclusions appropriate?*
- *Is the lack of exclusion analyses a threat to validity?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- Statistical risk model with 170 potential risk factors and 12 episode specific subtypes
- The developer states its goal was to complete an explanatory model rather than achieve parsimony. Therefore, no formal analysis was conducted to select risk factors.
- The developer used logistic regression to model the probability of having at least one PAC during the episode, including all covariates collected as risk factors.
- The developer used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample), and 2) the test data set (20% of sample).
- The developer did not include SDS variables in the model because they were not available in its dataset.
- Additional details on the conceptual methods, clinical methods, and statistical methods for the risk model are provided in the testing attachment.
- The developer reports the following discrimination and calibration statistics:

Statistical Risk Model Discrimination Statistics

Sample	Accuracy (%)*	AUC (C-Statistic)
Development	64.7%	0.700
Validation	65.8%	0.720

Statistical Risk Model Calibration Statistics

Sample	Chi Square	Degrees of Freedom	p-value
Development	30.4972	8	<0.0001
Validation	8.1555	8	0.4184

- According the developer, the results indicate the model has good predictive power.
  - The developer states the C-statistic on the development and validation samples (0.700 and 0.720, respectively) is considered to have good discriminatory power
  - The developer states the accuracy values show that the testing model correctly predicts whether an episode had or did not have a PAC approximately 65% of the time, above what would be expected if the predictions were made at random.

Questions for the Committee:

- *Is the risk adjustment strategy appropriate?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer calculated a risk-standardized PAC rate for each provider. Specifically, it calculated the expected number of episodes with PACs for each provider's panel of attributed patients.

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

Facilities:

PAC Rates	Minimum # Episodes Per Facility	
	>=10	>=125 (Reliability >0.7)
Unadjusted		
Median (IQR)	63% (58%, 69%)	63% (58%, 68%)
Range	27%-100%	48%-77%
Adjusted (RSPR)*		
Median (IQR)	63% (57%,69%)	63% (57%, 67%)
Range	30%-91%	50%-78%

Physicians:

PAC Rates	Minimum # Episodes Per Physician
	>=10 (Reliability >0.7)
Unadjusted	
Median (IQR)	60% (43%, 79%)
Range	0%-100%
Adjusted (RSPR)*	
Median (IQR)	58% (44%, 70%)
Range	0%-100%

*RSPR = Risk Standardized PAC Rate

- The developer concludes that the variation in risk-standardized PAC rates suggests meaningful differences in performance among providers that manage pneumonia at both the level of facilities and physicians.

Questions for the Committee:

- *Does this methodology demonstrate meaningful differences **across** measured entities?*
- *Does this measure identify meaningful differences in quality among facilities? among physicians?*
- *Should the Committee recommend the developer specify a minimum sample size directly in the specifications?*

2b6. Comparability of data sources/methods:
Not applicable

2b7. Missing Data

- The developer states "Not applicable," but elsewhere in the submission states if gender is missing the patient is eliminated and "If data is missing, the case is deleted from both the numerator and denominator."

Question for the Committee:

- *Does the Committee wish to clarify the issue of missing data with the developer?*

Guidance from Validity Algorithm: 1 → 2 → 3 → 4 → 5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Specifications are consistent with evidence in 1a.

Face validity. PAC analysis presented to physicians in 3 different healthcare systems, as well as medical directors from the employer coalition that provided the dataset to run the analyses. The developer stated the face validity testing was positive, and the physicians to whom the analyses were shown concurred with the construction of the PACs and their definitions. The developer did not report specifically on a face validity assessment at the level of the measure score, as required by NQF.

The developer presents significantly expanded face validity in the current submission (i.e., number of individuals and types of entities). Although a systematic assessment is not presented, the developer states it has not had results rejected in whole or part. Based on the reference to "results," NQF staff infer this indicates face validity at the performance score level.

The developer reports assessing face validity through application as follows:

Claims data analyses to physicians in different healthcare systems
Pennsylvania Employee Benefits Trust Fund, local provider groups and hospitals, Horizon Blue Cross Blue Shield of NJ, and physicians and health systems
Analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:
Vermont Payment Reform Commission
Maine Health Management Coalition
WellPoint / Anthem CT
NY State Medicaid
CT Medicaid
CO All-payer Claims Database, Center for Improving Value in Health care
Validity testing results:
The developer states it has not had results "rejected in whole or part," which NQF staff infers as face validity at the performance score level.
**The validity was tested on the measure and appears to be consistent with the target population.
**are specs c/w evidence - YES
**I have concerns about whether the outcomes are truly PACs. Events such as legionella and aspiration pneumonia re considered PACs. Wouldn't these be the cause of admission? Also, if a patient intubated or in septic shock at presentation, how can these events be considered PACs?
**Developer lists validity challenges related to risk adjustment "3. Claims data are messy with incomplete or incorrect diagnosis codes being used and there is limited clinical information, some of which can only be obtained through patient chart reviews such as information on body weight, BP, smoking status etc." However there are codes for tobacco use and obesity in ICD9 and ICD10 codes, both of which I verified were included in the risk factor list in the Excel document. Unclear on the challenge referenced by developer
** I am dubious that simply excluding ventilator associated pneumonia, fungal pneumonia and OIs will exclude HCAP. Why aren't those with gram negative or multidrug resistant pneumonia excluded?
I am unclear as to how PACs present are admission are handled. For example, I believe that a patient presenting to a hospital with sepsis due to pneumonia would be considered to have a PAC.
I was unable to review the actual codes. I worry about different degrees of preventability. I also suspect that patients treated as an out-patient will have lower risk of PACs – simply due to less testing that might identify what has been called a PAC – for example, an electrolyte disturbance is of dubious clinical relevance and unlikely to be found in a nonhospitalized patient.
** I am unsure as to how present on admission (or presentation) are handled. For example, of the more common PACs – many are likely present at the time of diagnosis. Here are all those present in at least 1% (table inserted)
-I suspect that a significant number of the PACS above were present at the time of pneumonia diagnosis or before (?Bronchiectasis). I highlighted those I suspect most likely to be present very early. Of the Type 2 PACS, I am unclear as to how they were determined to be Safety failures. For example, how does the developer know that aspiration pneumonia isn't, in fact, the initial pneumonia, rather than a complication? Legionnaire's disease is very likely Legionella pneumonia as a presenting complaint – not a complication. Same with MRSA.

2a2. Reliability Testing
Comments:
**Face validity. PAC analysis presented to physicians in 3 different healthcare systems, as well as medical directors from the employer coalition that provided the dataset to run the analyses. The developer stated the face validity testing was positive, and the physicians to whom the analyses were shown concurred with the construction of the PACs and their definitions. The developer did not report specifically on a face validity assessment at the level of the measure score, as required by NQF.
The developer presents significantly expanded face validity in the current submission (i.e., number of individuals and types of entities). Although a systematic assessment is not presented, the developer states it has not had results rejected in whole or part. Based on the reference to "results," NQF staff infer this indicates face validity at the performance score level.
The developer reports assessing face validity through application as follows:
Claims data analyses to physicians in different healthcare systems
Pennsylvania Employee Benefits Trust Fund, local provider groups and hospitals, Horizon Blue Cross Blue Shield of NJ, and physicians and health systems
Analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:
Vermont Payment Reform Commission
Maine Health Management Coalition
WellPoint / Anthem CT
NY State Medicaid

CT Medicaid
CO All-payer Claims Database, Center for Improving Value in Health Care
The developer states it has not had results "rejected in whole or part," which NQF staff infers as face validity at the performance score level.
**Validity was tested and would allow for conclusions regarding the quality of care. The score is an indicator of quality of care provided. Validity testing was done at the performance score level. The data to support the validity testing was one of the weaker elements of this measure summary.
**sufficient validity so conclusions re quality can be made? missing empirical validty testing.
**Insufficient based on the lack of a score.
**Consider stratification of the measure, does not require additional collection and would allow for a greater understanding of specific PACs requiring focus in terms of prevention of complications
** Reliability testing suggests a degradation in reliability if fewer than 125 patients included for facilities – only 18% of the facilities tested met this threshold. This does not appear to be a similar concern for provider level analyses. I am unsure why this is the case. These data also cover two years – if this is reported annually, might fewer than 10% of facilities have adequate samples?

2b2. Validity Testing
Comments:
**The developer reports the following denominator exclusions:
Patients < 8 years
Patients whose gender is missing
Patients who do not have continuous enrollment for the entire time window with the entity providing the data
Patients whose pneumonia episode time window extends outside the dataset time period
The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type)
"Claims" are excluded from the pneumonia measure if they are considered not relevant to pneumonia care
The developer did not complete formal exclusion testing or formal analysis on the impact of exclusions on performance scores as requested by NQF.
According the developer, the results indicate the model has good predictive power.
The developer states the C-statistic on the development and validation samples (0.700 and 0.720, respectively) is considered to have good discriminatory power
The developer states the accuracy values show that the testing model correctly predicts whether an episode had or did not have a PAC approximately 65% of the time, above what would be expected if the predictions were made at random.
The developer calculated a risk-standardized PAC rate for each provider. Specifically, it calculated the expected number of episodes with PACs for each provider's panel of attributed patients.
The developer concludes that the variation in risk-standardized PAC rates suggests meaningful differences in performance among providers that manage pneumonia at both the level of facilities and physicians.
Missing Data: The developer states "Not applicable," but elsewhere in the submission states if gender is missing the patient is eliminated and "If data is missing, the case is deleted from both the numerator and denominator."
**The data could support differences in outcomes and performance by facility. It did not describe any testing that would have led to conclusions on validity of the measure related to sub populations.
**exclusions appropriate? yes
lack of exclusion analysis a threat? yes
methodology demonstrate meaningful differences across entities? yes, R> 0.7
committee recommend minimum sample size? yes
yes - wish to clarify issue regarding missing data.
** Only face validity was measured.
**2b3. Yes, exclusions look appropriate.
Yes, lack of exclusion analysis is a threat to validity, especially given the exclusion of outliers.
** Patients with discontinuous enrollment for the time window are excluded. Claims are excluded if not relevant to pneumonia care – how is this determined.
2b4. Risk adjustment explains small part of variation. It seems to meet minimal criteria. May not be explanatory because the relevant data is missing from the data store.
Yes, variables are will described.
** AUC is 0.72 in validation cohort. Calibration p=0.42. Risk adjusting does not appreciable change the point or range estimates. Individual changes from crude to risk adjusted performance are not provided. If there are two classes of PACs, why is a single risk adjustment used? Seems like risks for safety failures will be different than those for a readmission.
2b5. Yes, especially across physicians.
It is difficult to know whether the differences are due to quality differences or differences in other factors.

As above, do think that there is a minimum sample size for facilities. I would be interested to see if the same range of outcomes would occur for physicians is the sample size was increased.
** I would be interested in seeing which PACs or HACs drive the range in performance.

2b6. NA
2b7. Sure
2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**The developer uses a beta-binomial method to demonstrate that the measure sufficiently differentiates performance among providers.
The developer states there is not a clear cut-off for minimum reliability level, but citing Adams (2009), notes values above 0.7 are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians or pairs of facilities.
The developer reports the following reliability testing results for facilities:
Facilities with at least 10 episodes: 25% of scores above the minimum accepted level of 0.7.
Facilities with at least 80 patients: median reliability score was 0.7
Facilities that exceeded 125 attributed patients: reliability scores consistently above 0.70
The developer concluded that for facilities with a high number of episodes, the pneumonia PAC measure sufficiently differentiates facility performance
The developer reports the following reliability testing results for physicians:
Physicians with at least 10 episodes: more than 75% reliability scores of 0.78 or higher; 10% of physicians in the sample had scores below 0.70.
The developer concluded the measure differentiates performance across a wide range of sample sizes.
**Reliability testing was conducted with an adequate sample size to generalize the results for widespread implementation.
**sufficient reliability so that differences in performance can be identified? Yes
recommend minimum sample size? yes - 80/10 vs 125/11
**Relaibility: 2 Low
The fact that a facility (which will be main unit of measure) requires 125 episodes for reliability of > 0.7 is very problematic. What percentage of hospitals would have this many episodes in a month/quarter? Can the developer explain why the relaibiltiy for physicians requires so many fewer patients?
I do believe we should recommend a minimum sample size of 125.
**Testing type, level of analysis and results there are no questions re: score level data
** Please clarify the missing data issue – especially related to gender and how this will be handled.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims.
- There are no fees associated. The measure specifications along with the metadata file are available without fees.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**The developer reports:

PAC measures are in use with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system. Included in different pilot site implementations:
BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers
PAC measures are incorporated by the following organizations in their bundled payment programs, as well as, process improvement activities and for practice re-engineering:
BCBSSC – for CABG and PCI programs
Horizon BCBSNJ– for CHF and CABG programs
BCBSNC
PEBTF in PA
Rates have been reported for PACs for the following organizations:
Vermont Payment Reform
All data elements are in defined fields in electronic claims.
There are no fees associated. The measure specifications along with the metadata file are available without fees.
**The measure as described in the methodology is feasible and the data elements are available electronically and through administrative data.
**feasability demonstrated
**High
**Elements are available, however with this being a claims based measure, not all elements may be captured in claims and require access to clinical information (e.g., vital signs- BP, O2 sats etc.) for assessment of risk factors
** Comes from claims data.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

Current Use:

- Payment Program:
  - Blue Cross Blue Shield of North Carolina
  - Horizon Blue Cross Blue Shield of New Jersey
  - Pennsylvania Employee Benefits Trust Fund
- Quality Improvement (internal to the specific organization)
  - Blue Cross Blue Shield of North Carolina

The developer reports:

- PAC measures are in use with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system. Included in different pilot site implementations:
  - •BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
  - •BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers
- PAC measures are incorporated by the following organizations in their bundled payment programs, as well as, process improvement activities and for practice re-engineering:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

- Rates have been reported for PACs for the following organizations:
    - Vermont Payment Reform
    - Maine Health Management Coalition
    - WellPoint / Anthem CT
    - NY State Medicaid
    - CT Medicaid
    - CO All-payer Claims Database, Center for Improving Value in Health Care
- Companies are leveraging PAC measures to create analytics and software for customers, including:
    - HealthQx
    - Aver Informatics
    - McKesson
    - TriZetto.
    - FairHealth
- PAC measures have been adopted as category 1 quality measures for New York State DSRIP (Delivery System Reform Incentive Payment) project for Medicaid Redesign.

Planned Use:

- Public reporting
- Professional certification or recognition programs
- Quality improvement with benchmarking
- The developer reports it is working with not-for-profit and for-profit organizations to provide algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:
    - Fair Health
    - CastLight
    - MA APCD (Massachusetts All Payers Claims Database) Council
    - Maryland Health Care Cost Commission

Improvement results

- The developer does not share improvement results, since implementation is too recent to provide valid comparisons. In addition, measure specifications have changed since the initial endorsement, so year-over-year comparisons would not be valid.

Unexpected findings (positive or negative) during implementation

- No information provided by developer.

Potential harms

- The developer notes no unintended consequences experience, but acknowledges the potential for:
    - Under-coding of PACs in the claim stream
    - Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

Feedback :

- MAP reviewed the measure in 2015 Hospital In-Patient Quality Reporting Program. MAP decided to review the measure as a fully developed measure and conditionally support its use in the IQR program pending successful testing and NQF endorsement of the measure with an expanded population.

***Questions for the Committee**:*

- *Gaming by under coding claims is identified as a potential concern, do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**PAC measures are in use with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system. Included in different pilot site implementations:
BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers
PAC measures are incorporated by the following organizations in their bundled payment programs, as well as, process improvement activities and for practice re-engineering:
BCBSSC – for CABG and PCI programs
Horizon BCBSNJ– for CHF and CABG programs
BCBSNC
PEBTF in PA
Rates have been reported for PACs for the following organizations:
Vermont Payment Reform
Maine Health Management Coalition
WellPoint / Anthem CT
NY State Medicaid
CT Medicaid
CO All-payer Claims Database, Center for Improving Value in Health Care
Companies are leveraging PAC measures to create analytics and software for customers, including:
HealthQx
Aver Informatics
McKesson
TriZetto.
FairHealth
PAC measures have been adopted as category 1 quality measures for New York State DSRIP (Delivery System Reform Incentive Payment) project for Medicaid Redesign.
Planned Use:
Public reporting
Professional certification or recognition programs
Quality improvement with benchmarking
The developer reports it is working with not-for-profit and for-profit organizations to provide algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:
Fair Health
CastLight
MA APCD (Massachusetts All Payers Claims Database) Council
Maryland Health Care Cost Commission

Improvement results
The developer does not share improvement results, since implementation is too recent to provide valid comparisons. In addition, measure specifications have changed since the initial endorsement, so year-over-year comparisons would not be valid.
Unexpected findings (positive or negative) during implementation
No information provided by developer.
Potential harms
The developer notes no unintended consequences experience, but acknowledges the potential for:
Under-coding of PACs in the claim stream
Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers
Feedback :
MAP reviewed the measure in 2015 Hospital In-Patient Quality Reporting Program. MAP decided to review the measure as a fully developed measure and conditionally support its use in the IQR program pending successful testing and NQF endorsement

**The measure demonstrates usability though at a population level. No unintended consequences noted. Implementation should result in the goal of high-quality, efficient healthcare.
**do benefits of measure outweigh any potential unintended consequences - needs committee discussion
**I think it is reasonable to think that hospitals will look hard at whether to code complications once they know that they are being graded on their rates. It would be good to see improvements in quality brought about by this measure to better assess whether the benefit outweighs risks.
**Measure is certainly useful and can be used, but could benefit for clarifications for consistency in implementation or at a minimum for understanding of how data are generated. Also, stratification by PAC categories (at least) and then looking to see if there are any disparities in the data would be helpful in targeting improvement
** Largely used by private payers. No data on any improvement. Without limitations in minimal cohort size, I worry about penalties for low point estimates without any appreciation of the precision of the measure in this group.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0506: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
- 1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

Harmonization

- The developer states not all measures specifications are completed harmonized

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0708

Measure Title: Proportion of Patients with Pneumonia that have a Potentially Avoidable Complication (during the episode time window)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.

- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Potentially Avoidable Complications

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Pneumonia is a leading cause of mortality and morbidity in the US population (CDC 2015). There may be up to 5 to 6 million cases of community-acquired pneumonia (CAP) diagnosed annually in the United States, accounting for approximately 1 million hospitalizations and approximately 10 million physician visits (Aliberti 2008). Hospitalizations and readmissions due to community-acquired pneumonia are common and may not always be warranted (AHRQ 2015). Pneumonia has long been subject to quality improvement efforts and was often used as a marker for hospital quality (Meehan 2001). Clear guidelines have been published outlining criteria for admissions (PSI criteria, CURB-65 criteria). Adherence to guidelines results in better survival and more favorable outcomes in patients with pneumonia (Niederman 2001, Guann-Ming Chang 2011), perhaps due to shifting of care into outpatient setting and due to decrease in inpatient complications (CarratalàJ 2005, Martinez 2009).

Potentially avoidable complications (PACs) are the unwarranted health outcomes that this measure addresses. PACs are both directly and indirectly related to healthcare services provided (or not provided) during the episode time window (de Brantes 2009) (James 2013). PACs may occur due to errors in omission or commission. Errors of omission in a patient with pneumonia could be due to failure of hospitals and / or physicians to give the right antibiotics at the right time, failure to check oxygen saturation, do a mental status assessment or note rapid deterioration in vital signs leading to progression of sepsis, respiratory insufficiency culminating in respiratory failure and even death (AHRQ 2002). Failure to establish or implement patient safety processes such as evidence-based protocols to prevent deep vein thrombosis, pressure ulcers, GI hemorrhage from stress of hospitalization etc. may lead to other harms while the patient is in the hospital. In addition, errors of commission such as line sepsis, hospital acquired infections etc. could occur due to poor processes followed during insertion of central lines and other invasive procedures during the hospitalizations (Provonost 2010). Lack of care coordination, poor discharge planning and poor arrangements of patient follow-up could lead to unnecessary ER visits, readmissions and gaps in care leading to increased morbidity and repeat hospitalizations (Weaver 2013). All these adverse events are aggregated together in a single measure of provider performance -- to study the overall rate of PACs.

Scenario 1:

Timely administration of Antibiotics in a Pneumonia Patient

↓

Respiratory Insufficiency / Sepsis / Hospital-Acquired Infection Avoided

↓

PAC avoided

Scenario 2:

Adult patient with Pneumonia

↓

Hospital/physician fails to carry out safe processes (error in commission/omission)

↓

Patient suffers complication stemming from hospital/physician potentially avoidable error

↓

Patient remains in hospital for treatment of PAC

OR

Patient readmitted to hospital with 1+ Potentially avoidable complication

OR

Patient closely followed up in the post-discharge period due to consequences of a PAC

Other scenarios:

There are a wide variety of other ways to reduce PACs at the healthcare provider level for an individual with pneumonia. A few additional examples of better processes of care leading to reduced PACs are given below:

1. Frequent hand-washing reduce hospital acquired infections (WHO 2007)
2. Carefully implemented protocols lead to reduced line sepsis (Pronovost 2010)
3. Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
4. DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)
5. Frequent change in position of AMI patients in the CCU avoids pressure sores (Sullivan 2013)

Better processes of care create an atmosphere of proactive management, consistency in care and standardized care patterns. These processes significantly reduce occurrence of potentially avoidable complications (PACs) in all settings – while the patient is in the hospital and post-discharge (Wachter 2013). Additionally, a study from the Boston Medical Center, Boston MA, demonstrated that although one in five hospitalizations are complicated by post-discharge adverse events, development of a strong discharge services program for patients admitted for medical conditions reduced hospital utilization within 30 days of discharge.

Umscheid et al used 2002 estimates of hospital-acquired infections (HAI) and determined the range of HAI risk reductions from US studies. They report that 18%-82% of blood-stream infections, 46%-55% of ventilator associated pneumonia, 17% - 69% of urinary tract infections and 26%-54% of surgical site infections are preventable. Healy et al analyzed complications in hospitalized surgical patients and reported that between 39% and 61% of major complications (wound infections, pneumonia, urinary tract infections, arrhythmias, respiratory failure, gastrointestinal complications, deep vein thrombosis) and about an equal percent of minor complications could have been avoided. The National Pressure Ulcer Advisory Panel (NPUAP) reported in 2001 that pressure ulcer prevention programs had reported 50% or greater reductions in facility-acquired pressure ulcers (Cuddigan 2001). Similarly, appropriate prophylaxis could reduce the risk of venous thromboembolism by 45% in acutely ill medical patients (Leizorowicz 2004), and a randomized placebo-controlled trial found a 50% reduction in thromboembolic events with extended pharmacologic prophylaxis (Hull 2007). Adequate evidence-based treatment protocols in preventing contrast nephropathy and adequate

drug dosing have demonstrated a risk reduction between 52% and 90% in the incidence of acute renal failure in patients in the intensive care unit (Singri 2003). Additionally, use of electronic medical systems has demonstrated that in a sample hospital that used prompts for protocols for nursing care, infection rates dropped 88%, bedsores were reduced and compliance to guidelines for care of patients on ventilator increased by 77% (Landro 2009).

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient, as well as for the health plan with whom the patient is a member (de Brantes 2009).

References:

1) Centers for Disease Control and Prevention: http://www.cdc.gov/nchs/fastats/deaths.htm - Accessed Dec 6th 2015.
2) Aliberti S, Amir A, Peyrani P et. Al. Incidence, Etiology, Timing, and Risk Factors for Clinical Failure in Hospitalized Patients with Community-Acquired Pneumonia. CHEST 2008; 134: 955-962.
3) Agency for Health Research and Quality 2013: http://archive.ahrq.gov/research/findings/factsheets/pneumonia/issue7/ - Accessed Dec 6th 2015.
4) Meehan TP, Weingarten SR, Holmboe ES, et al. A statewide initiative to improve the care of hospitalized pneumonia patients: the Connecticut Pneumonia Pathway Project. Am J Med. 2001;111(3): 203-210.
5) PSI criteria: Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336(4):243.
6) CURB-65 criteria: Guidelines for the management of community-acquired pneumonia in adults admitted to hospital. The British Thoracic Society. Br J Hosp Med. 1993; 49(5):346.
7) Guann-Ming Chang, MD MS and Yu-Chi Tung, PhD. Factors Associated with Pneumonia Outcomes: A Nationwide Population-Based Study over the 1997–2008 Period J Gen Intern Med 2011; 27(5):527–33 DOI: 10.1007/s11606-011-1932-1
8) Niederman MS, Mandell LA, Anzueto, et al. Guidelines for the management of adults with community-acquired pneumonia: diagnosis, assessment of severity, antimicrobial therapy, and prevention. Am J Respir Crit Care Med 2001; 163:1730-1754.
9) CarratalàJ, Fernández-SabéN, Ortega L, CastellsaguéX, Rosón B, Dorca J, Fernández-Agüera A, Verdaguer R, Martínez J, Manresa F, Gudiol F. Outpatient care compared with hospitalization for community-acquired pneumonia: a randomized trial in low-risk patients. Ann Intern Med. 2005;142(3):165.
10) Martínez R, Reyes S, Lorenzo MJ, Menéndez R. Impact of Guidelines on Outcome: the evidence. Semin Respir Crit Care Med 2009; 30(2): 172.
11) de Brantes, François M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. "Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model." NEJM (2009) 361:1033 (Perspective)
12) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care. " J Patient Saf 9.3 (2013): 122-128.
13) Stanton MW. Improving Treatment Decisions for Patients with Community-Acquired Pneumonia. Rockville (MD): AHRQ; 2002. Research in Action Issue 7. AHRQ Pub. No. 02-0033.
14) P. J. Pronovost, G. A. Goeschel, E. Colantuoni et al., "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," BMJ 340:c309 (Feb. 4, 2010): Web.
15) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." Annals of Internal Medicine 158.5_Part_2 (2013): 369-75. Web.
16) The Joint Commission, Aide Joint Commission International, and WHO. "Improved Hand Hygiene to Prevent Health Care-Associated Infections." World Health Organization. WHO Collaborating Centre for Patient Safety Solutions (May 2007). Web.

17) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." Annals of Internal Medicine 158.5_Part_2 (2013): 365-69. Web.
18) Sullivan N, Schoelles KM. "Preventing in-facility pressure ulcers as a patient safety strategy. A systematic review." Ann Intern Med 158 (2013): 410-6.
19) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." Annals of Internal Medicine 158.5_Part_1 (2013): 350-52. Web.
20) Umscheid CA, Mitchell MD, Agarwal R, et al, Mortality from Reasonably-Preventable Hospital-Acquired Infections, (Philadelphia Penn Center for Evidence-based Practice Advisory, 2008). Available at http://www.shea-online.org/Assets/files/0408_Penn_Study.pdf Accessed April 27, 2009.
21) Ranji SR, Shetty K, Posley KA, et. Al, Prevention of Healthcare-Associated Infections. Vol 6 of Shojania KG, McDonalds KM, Wachter RM, Owens DK, editors. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Technical Review 9 (Prepared by the Stanford University – UCSF Evidence-based Practice Center under Contract No. 290-02-0017). AHRQ Publication No. 04(07)-0051-6. Rockville, MD: Agency for Health Research and Quality. January 2007.
22) Healey M, Shackford SR, Osler TM, et al. "Complications in Surgical Patients," Archives of Surgery 137, no. 5 (2002): 611-618.
23) Cuddigan J, Berlowitz DR, Ayello EA. Pressure Ulcers in America: Prevalence, Incidence, and Implications for the Future: An Executive Summary of the National Pressure Ulcer Advisory Panel Monograph. Advances in Skin and Wound Care 2001, 14(4) 208-215.
24) Leizorowicz A, Cohen AT, Turpie AF, et al: Randomized placebo-controlled trial of dalteparin for the prevention of venous thromboembolism in acutely ill medical patients. Circulation 2004; 110:874-9.
25) Hull RD, Schellong SM, Tapson VF, et al. Extended-duration enoxaparin prophylaxis in high-risk medical patients. Abstract presented at International Society on Thrombosis and Hemostasis, July 8-11, 2007, Geneva, Switzerland.
26) Singri N. "Acute Renal Failure," JAMA 289, no. 6 (2003): 747-751.
27) Landro L. An Affordable Fix for Modernizing Medical Records. http://online.wsj.com/article/SB124104350516570503.html Accessed April 30th 2009.
28) de Brantes F, D'Andrea G, Rosenthal MB. "Should health care come with a warranty?" Health Aff (Millwood) 28 (2009): w678-w687.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Rationale:

A significant portion of waste in today's healthcare system is due to "care defects" – errors, avoidable hospitalizations, and other process failures that cause patients to incur unnecessary services and some harm. For example, a report by the Agency of Health Care Research and Quality (AHRQ) highlighted the fact that in 2008, 1 out of every 10 hospital-stays could have been prevented (Stranges 2008), and for Medicare beneficiaries one in five admissions were for a potentially preventable condition. To improve accountability in the delivery of medical care, AHRQ has developed a list of patient safety indicators (PSIs) to identify potential harms to patients (AHRQ 2008). Additionally, the Centers for Medicare and Medicaid Services (CMS) have taken a "Six Sigma" approach and defined Hospital Acquired Conditions (HACs) and "never events" that should almost never occur and are applying financial penalties when these events do occur (CMS 2012). The Potentially avoidable complications (PAC) measure goes beyond the AHRQ PSIs and the CMS HACs and creates a single comprehensive measure that measures all-cause harms for a patient with the index condition.

Potentially avoidable complications may occur anytime during the course of a patient's illness, especially if pneumonia patients get hospitalized. Once hospitalizations occur, the index stay itself may have a potentially

avoidable complication (PAC) or patients may develop a PAC during the 30-day post-discharge period. PACs lead to significant variability in outcomes including prolonged length of stay, readmissions and emergency room visits, all indicating poor outcomes that harm the patient, cause payers to incur unnecessary costs and could be improved by providers (de Brantes 2011).

A comprehensive measure of "all-cause" harms tracks all adverse events in one measure and creates a complete picture for the provider's performance. It looks at all care defects such as readmissions, emergency room visits, adverse events due to errors of omission or commission. It looks at complications that are directly related to the index condition, and also those due to patient safety failures. As such, the measure provides clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer. Performance results provide summary PAC rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider network, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs. It is also known that by holding providers accountable for occurrence and costs of PACs, a built-in warranty is created around care of the index condition (de Brantes 2009) (Curry 2011) and it also creates incentives to reduce readmissions (Mittler 2013).

Further, as a comprehensive outcome measure, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues where they found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating about the quality of care to consumers (Hibbard 2011). It helped consumers make more rational choices among providers, not equating high prices to high quality, but in fact if high prices were stemming from high PACs, then it may indicate more waste and more harm.

References:

1) Stranges, Elizabeth, MS, and Carol Stocks, RN, MHSA. Potentially Preventable Hospitalizations for Acute and Chronic Conditions, 2008. Rep. Vol. 99. N.p.: Agency for Healthcare Research and Quality. Web.
2) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. ''AHRQ Quality Indicators. Prevention Safety Indicators (PSIs): Technical Specifications, Version 3.2'': http://www.qualityindicators.ahrq.gov/modules/PSI_TechSpec.aspx
3) Centers for Medicare and Medicaid Services. "Hospital Acquired Conditions." CMS.gov. Centers for Medicare and Medicaid Services, 2012. Web.

4) de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
5) de Brantes F, D'Andrea G, Rosenthal MB. "Should health care come with a warranty?" Health Aff (Millwood) 28 (2009): w678-w687.
6) Curry L.A. , Spatz E., Cherlin E. et al., "What Distinguishes Top-Performing Hospitals in Acute Myocardial Infarction Mortality Rates?" Annals of Internal Medicine 154.6 (March 15, 2011): 384–90.
7) Mittler, Jessica N., Jennifer L. O'hora, Jillian B. Harvey, Matthew J. Press, Kevin G. Volpp, and Dennis P. Scanlon. "Turning Readmission Reduction Policies into Results: Some Lessons from a Multistate Initiative to Reduce Readmissions." Population Health Management 16.4 (2013): 255-60. Web.
8) Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. Health Affairs 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0708_PNE_Evidence_Attachment_HCI3.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of potentially avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, hospital, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 60% of its plan members with pneumonia incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and

implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow-up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events" and non-payment for certain readmissions) and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x
2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.
3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.
4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt
5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.
6) Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf
7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.
8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html
9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.
10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/
11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.
12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.
13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.
14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.
15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.
16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.
17) Leibson CL1, et al. "Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort." Med Care 46.2 (2008):127-32.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 13,228 episodes of pneumonia with over 54% of episodes having one or more potentially avoidable complications. The enclosed workbook entitled NQF_PNE_all_codes_risk_adjustment_12_14_15.xls lists the types of PACs and their frequency as calculated in this database. The PAC overview tab shows the PAC rates in various settings across all episodes and the PAC drill down graph gives the details for each individual PAC type.

Over 27% of Community Acquired Pneumonia patients were managed in an inpatient setting, and another 30% had an outpatient facility encounter. The remaining patients were managed completely in doctor's offices and in the community setting. 54% of all pneumonia episodes had a PAC, 20% of them being observed in an inpatient setting, 13% in an outpatient setting and 50% in a professional setting. Majority of the PACs incurred were directly related to the pneumonia (type 1 PACs: 91%) such as respiratory insufficiency (23%), other lung complications (19%), fluid and electrolyte problems (14%), respiratory failure (12%) and sepsis (9%). Some of these patients also had Type 2 PACs related to patient comorbidities or patient safety failures. Overall 21% of patients had a type 2 PAC, such as urinary tract infection (4.5%), diabetes poor control (4%), aspiration pneumonia (3%), deep vein thrombosis (2.9%) and pulmonary embolism (2.5%). Four percent of all pneumonia patients had a readmission. The primary cause for readmissions and emergency room visits during the 30-day post-discharge period was due to respiratory failure, a repeat pneumonia, or sepsis.

Health plan PAC scores help look at network performance and helps plans target improvement opportunities through patient engagement tools, nurse help lines or by network management.

Provider level PAC scores were also calculated. Pneumonia episodes were attributed to the treating facilities in cases where the patients were hospitalized, and in a second attribution exercise to physicians who were primarily responsible for managing the pneumonia (those with the maximum number of E&M services).

Because providers with small volumes may provide unreliable estimates, we excluded any providers with fewer than 10 attributed episodes prior to the calculations. In the current database, 82 facilities and 170 physicians had cared for at least 10 patients during the analysis time period. Performance scores (PAC rates) for these facilities and providers are summarized in the following table:

Facility Unadjusted PAC Rates (n=82):
- Median (IQR): 63% (58%, 69%)
- Range: 27% - 100%

Facility Risk-Standardized PAC Rates (RSPR):
- Median (IQR): 63% (57%, 69%)
- Range: 30% - 91%

Physician Unadjusted PAC Rates (n=170):
- Median (IQR): 60% (43%, 79%)
- Range: 0% - 100%

Physician Risk-Standardized PAC Rates (RSPR):
- Median (IQR): 58% (44%, 70%)
- Range: 0% - 100%

Please refer to the NQF_PNE_all_codes_risk_adjustment_12_14_15.xls workbook under the "Facility RSPR & Reliability" and "Physician RSPR & Reliability" tabs to see specific results for each facility and physician respectively.

The variation in risk-standardized PAC rates across providers suggests there the measure identifies meaningful differences in performance among providers that manage pneumonia at both the level of facilities and physicians

The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient, as well as for the health plan with whom the patient is a member.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Pneumonia is a leading cause of mortality and morbidity in the US population (CDC 2015). Hospitalizations and readmissions due to community-acquired pneumonia are common and may not always be warranted (AHRQ 2015). Pneumonia has long been subject to quality improvement efforts and was often used as a marker for hospital quality (Meehan 2001). It has been shown that weekend admissions and decreased hospital reimbursements may lead to worse outcomes suggesting opportunity for improvement by implementing better processes in care (Guann-Ming Chang 2011). Clear guidelines have been published outlining criteria for admissions (PSI criteria, CURB-65 criteria, Niederman 2001). Adherence to guidelines results in better outcomes (Martinez 2009).

With the publication of clear-cut guidelines for selection of patients for outpatient management; there have been reports of better survival and more favorable outcomes in patients with pneumonia (CarratalàJ 2005), perhaps due to shifting of care into outpatient setting and due to decrease in inpatient complications. A recent report however suggested that the perceived improvement in pneumonia outcomes may be an artifact of secular trends in changes in documentation and coding towards sepsis and respiratory failure (Lindenauer 2012), suggesting an ongoing performance gap in the care of pneumonia patients.

Additionally, potentially avoidable complications may occur anytime during the course of a patient's illness, especially if pneumonia patients get hospitalized. Once hospitalizations occur, the index stay itself may have a potentially avoidable complication (PAC) or patients may develop a PAC during the 30-day post-discharge period. PACs lead to significant variability in outcomes including prolonged length of stay, readmissions and emergency room visits, all indicating poor outcomes that harm the patient, cause payers to incur unnecessary costs and could be improved by providers (de Brantes 2011).

Readmissions constitute an important part of the PAC measure. Two-thirds of eligible hospitals had readmission rates that were higher than that predicted by the CMS model, highlighting the continued need for better care coordination across providers and the community to prevent PACs (Dharamrajan 2013). Importantly, recent studies have shown that high performing hospitals had fewer readmissions within 30 days for all common diagnoses, suggesting possible benefits of adopting strategies to reduce readmissions globally (Jack 2009, Dharmarajan 2013).

While PACs may not be completely eliminated, identifying their magnitude and understanding their causality, in particular for the most frequent or the most expensive, could lead to improving patient outcomes (de Brantes 2008) (de Brantes 2009).

References

1) Centers for Disease Control and Prevention: http://www.cdc.gov/nchs/fastats/deaths.htm - Accessed Dec 6th 2015.
2) Agency for Health Research and Quality: http://archive.ahrq.gov/research/findings/factsheets/pneumonia/issue7/ - Accessed Dec 6th 2015.
3) Meehan TP, Weingarten SR, Holmboe ES, et al. A statewide initiative to improve the care of hospitalized pneumonia patients: the Connecticut Pneumonia Pathway Project. Am J Med. 2001;111(3): 203-210.
4) Guann-Ming Chang, MD MS and Yu-Chi Tung, PhD. Factors Associated with Pneumonia Outcomes: A Nationwide Population-Based Study over the 1997–2008 Period J Gen Intern Med 2011; 27(5):527–33 DOI: 10.1007/s11606-011-1932-1
5) PSI criteria: Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336(4):243.
6) CURB-65 criteria: Guidelines for the management of community-acquired pneumonia in adults admitted to hospital. The British Thoracic Society. Br J Hosp Med. 1993; 49(5):346.
7) Niederman MS, Mandell LA, Anzueto, et al. Guidelines for the management of adults with community-acquired pneumonia: diagnosis, assessment of severity, antimicrobial therapy, and prevention. Am J Respir Crit Care Med 2001; 163:1730-1754.
8) Martínez R, Reyes S, Lorenzo MJ, Menéndez R. Impact of Guidelines on Outcome: the evidence. Semin Respir Crit Care Med 2009; 30(2): 172.
9) CarratalàJ, Fernández-SabéN, Ortega L, CastellsaguéX, Rosón B, Dorca J, Fernández-Agüera A, Verdaguer R, Martínez J, Manresa F, Gudiol F. Outpatient care compared with hospitalization for community-acquired pneumonia: a randomized trial in low-risk patients. Ann Intern Med. 2005;142(3):165.
10) Peter K. Lindenauer, Tara Lagu, Meng-Shiou Shieh, Penelope S. Pekow, Michael B. Rothberg. Association of Diagnostic Coding With Trends in Hospitalizations and Mortality of Patients With Pneumonia, 2003-2009. JAMA. 2012;307(13):1405-1413
11) Dharmarajan, Kumar, et al. "Diagnoses and Timing of 30-Day Readmissions After Hospitalization for Heart Failure, Acute Myocardial Infarction, or Pneumonia." JAMA 309.4 (2013): 355. Web.
12) Jack BW, Chetty VK, Anthony D et al. "A Reengineered Hospital Discharge Program to decrease Rehospitalization: A randomized trial." Ann Int Med 150 (2009):178-187.
13) Dharmarajan, K., et al. "Hospital Readmission Performance and Patterns of Readmission: Retrospective Cohort Study of

Medicare Admissions." British Medical Journal 347 (2013): Medical Journal. Web.
14) de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
15) de Brantes F. and A. Rastogi, "Evidence-Informed Case Rates: Paying for Safer, More Reliable Care," The Commonwealth Fund 40, publ. 1146 (2008): 1-14.
16) de Brantes, François M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. "Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model." NEJM (2009): 361:1033 (Perspective)

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not Applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Various studies have documented health disparities between African-Americans and Caucasians hospitalized with common infectious diseases (Richardus 2001, Keppel 2007). Some studies demonstrated that African-American patients with community acquired pneumonia (CAP) do not receive timely antibiotic treatment or lag behind in other process measures such as smoking cessation counseling or pneumococcal or influenza immunizations (Fine 2002, Mortensen 2004, Bennett 1995, Hausmann 2009). However, a large multicenter study based on an analysis of over 40,000 patients with CAP managed over 5 years across 150 Veterans Health administration (VHA) hospitals demonstrated that African-Americans and Caucasians were equally likely to receive guideline-concordant antibiotics and experienced similar 30-day mortality when treated in medical wards. When admitted to the ICU, African Americans, in fact experienced a survival advantage with a lower 30-day mortality as well as shorter hospital LOS as compared to their Caucasians counterparts (Frei 2010).

References:
1) Richardus, Kunst AE: Black-white differences in infectious disease mortality in the United States. ?American Journal of Public Health 2001, 91(8):1251-1253. PubMed Abstract | Publisher Full Text | PubMed Central Full Text ?
2) Keppel KG: Ten largest racial and ethnic health disparities in the United States based on Healthy People 2010 Objectives. ?Am J Epidemiol 2007, 166(1):97-103. PubMed Abstract | Publisher Full Text ?
3) Fine JM, Fine MJ, Galusha D, Petrillo M, Meehan TP: Patient and hospital characteristics associated with recommended processes of care for elderly patients hospitalized with pneumonia: results from the Medicare quality indicator system pneumonia module. ?Arch Intern Med 2002, 162(7):827-833. PubMed Abstract | Publisher Full Text ?
4) Mortensen EM, Cornell J, Whittle J: Racial variations in processes of care for patients with community-acquired pneumonia. ?BMC Health Serv Res 2004, 4(1):20. PubMed Abstract | BioMed Central Full Text | PubMed Central Full Text ??
5) Bennett CL, Horner RD, Weinstein RA, Dickinson GM, DeHovitz JA, Cohn SE, Kessler HA, Jacobson J, Goetz MB, Simberkoff M, Pitrak D, George L, Gilman SC, Shapiro MF: Racial differences in care among hospitalized patients with Pneumocystis carinii pneumonia in Chicago, New York, Los Angeles, Miami, and Raleigh-Durham. ?Arch Intern Med 1995, 155(15):1586-1592. PubMed Abstract | Publisher Full Text ?
6) Hausmann LR, Ibrahim SA, Mehrotra A, Nsa W, Bratzler DW, Mor MK, Fine MJ: Racial and ethnic disparities in pneumonia treatment and mortality. ?Med Care 2009, 47(9):1009-1017. PubMed Abstract | Publisher Full Text ?
7) Christopher R Frei, Eric M Mortensen, Laurel A Copeland, Russell T Attridge, Mary Jo V Pugh234, Marcos I Restrepo, Antonio Anzueto, Brandy Nakashima and Michael J Fine. Disparities of Care for African-Americans and Caucasians with Community-Acquired Pneumonia: A Retrospective Cohort Study BMC Health Services Research 2010, 10:143 doi:10.1186/1472-6963-10-143. Published: 27 May 2010 http://www.biomedcentral.com/1472-6963/10/143 - Accessed Dec 8th 2015.

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality,

Severity of illness, Other

1c.2. **If Other:** Lot of variability in complication rates and care patterns

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Pneumonia is a leading cause of mortality and morbidity in the US population (CDC 2015). There may be up to 5 to 6 million cases of community-acquired pneumonia (CAP) diagnosed annually in the United States, accounting for approximately 1 million hospitalizations and approximately 10 million physician visits (Aliberti 2008). It is estimated that the annual cost of treating CAP in the United States is $12.2 billion (Colice 2004). Pneumonia is the second most common reason for hospitalization after childbirth (AHRQ - 2013). Hospitalizations and readmissions due to community-acquired pneumonia are common and may not always be warranted (AHRQ 2015). When hospitalizations do occur, they must be managed expeditiously and readmissions following discharge should be avoided (MedPac 2007, Meehan 2001).

The concept of preventable complications is well founded in medical literature. Hospital acquired conditions (HACs) have been defined by the Centers for Medicare and Medicaid (CMS) under the proposed rules for 2008 and 2009. Other potentially avoidable complications have been suggested by AHRQ as prevention safety indicators (PSIs). More broadly, potentially avoidable complications are rampant and programs are being set up in place to address them (Weaver 2013, Watcher 2013, Shekelle 2013). Umscheid et al (2008) used 2002 estimates of hospital-acquired infections (HAI) and determined the range of HAI risk reductions from US studies. They and others report that 18%-82% of blood-stream infections, 46%-55% of ventilator associated pneumonia, 17% - 69% of urinary tract infections and 26%-54% of surgical site infections are preventable (Ranji 2007). The National Pressure Ulcer Advisory Panel (NPUAP) reported in 2001 that pressure ulcer prevention programs had reported 50% or greater reductions in facility-acquired pressure ulcers (Cuddigan 2001). Similarly, appropriate prophylaxis could reduce the risk of venous thromboembolism by 45% in acutely ill medical patients (Leizorowicz 2004), and a recent study found a 50% reduction in thromboembolic events with extended pharmacologic prophylaxis (Hull 2007). Adequate evidence-based treatment protocols in preventing contrast nephropathy and adequate drug dosing have demonstrated a risk reduction between 52% and 90% in the incidence of acute renal failure in patients in the intensive care unit (Singri 2003). Additionally, use of hospital electronic medical systems has demonstrated that in a sample hospital that used prompts for protocols for nursing care, infection rates dropped 88%, bedsores were reduced and compliance to guidelines for care of patients on ventilator increased by 77% (Landro 2009).

Readmissions constitute an important part of the PAC measure. Readmissions are rampant and represent waste within the healthcare system (Dharmarajan 2013, Jiang 2006). The June 2007 MedPAC report to Congress on "Promoting Greater Efficiency in Medicare" highlighted the fact that in 2005, $12 billion were spent on potentially preventable readmissions alone within 30 days of discharge from the hospital. Another study by Jencks and colleagues found that roughly 19.6% of Medicare patients incurred re-hospitalizations within 30 days of discharge (Jencks 2009). Jack et al proposed a reengineered hospital discharge program to reduce readmission (Jack 2009). In Oct 2012, CMS initiated the Hospital Readmissions Reduction Program (HRRP), in an effort to reduce the readmission rate of Medicare patients (MedPAC 2013). If the readmission rate exceeds the expected readmissions rates, then financial penalties are imposed. The main impact of the HRRP has been to increase the efforts of hospitals to reduce readmissions. Two-thirds of eligible hospitals had readmission rates that were higher than that predicted by the CMS model, highlighting the continued need for better care coordination across providers and the community to prevent PACs (Jyont 2013). Subsequently, CMS reported a decline in readmissions ever since the penalties were imposed averting an estimated 150,000 hospitalization in just the fourth quarter of 2012 (MedPAC 2013).

Lewis et al suggest that a stratified approach targeting high impact conditions, using data to identify areas of opportunity and focused interventions with feedback loops could form a self-learning system that could avert "Triple Fail" events before they occur (Lewis 2013). While PACs may not be completely eliminated, identifying their magnitude and understanding their causality, in particular for the most frequent or the most expensive, could lead to improving patient outcomes (de Brantes 2008) (de Brantes 2009).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1) Centers for Disease Control and Prevention: http://www.cdc.gov/nchs/fastats/deaths.htm - Accessed Dec 6th 2015.

2) Aliberti S, Amir A, Peyrani P et. Al. Incidence, Etiology, Timing, and Risk Factors for Clinical Failure in Hospitalized Patients with Community-Acquired Pneumonia. CHEST 2008; 134: 955-962.

3) G. Colice et al. Treatment Costs of Community-Acquired Pneumonia in an Employed Population. Chest 2004; 125: 2140-2145.

4) Agency for Health Research and Quality 2013: http://archive.ahrq.gov/research/findings/factsheets/pneumonia/issue7/ - Accessed Dec 6th 2015.

5) MedPAC. Report to The Congress: Promoting Greater Efficiency in Medicare. 2007. Available at: http://www.medpac.gov/documents/jun07_EntireReport.pdf.

6) Meehan TP, Weingarten SR, Holmboe ES, et al. A statewide initiative to improve the care of hospitalized pneumonia patients: the Connecticut Pneumonia Pathway Project. Am J Med. 2001;111(3): 203-210.
7) Department of Health and Human Services Center for Medicare and Medicaid Services. Proposed Rules. 73(84):23547-23562. Available at: <http://edocket.access.gpo.gov/2008/pdf/08-1135.pdf>. Accessed October 29th 2009.
8) ECRI Institute. Special Advisory: List of CMS Hospital-Acquired Conditions Expanded under New Final Rule. 2008. Available at: <https://www.ecri.org/PatientSafety/HrcReports/Pages/CMS_Final_Rule_on_Hospital_Acquired_Conditions.aspx>. Accessed October 30th 2009.
9) Centers for Medicare and Medicaid Services. "Hospital Acquired Conditions." CMS.gov. Centers for Medicare and Medicaid Services, 2012. Web.
10) AHRQ. Quality Indicators: Guide to Patient Safety Indicators. 2007. Available at: <http://www.qualityindicators.ahrq.gov/downloads/psi/psi_guide_v31.pdf> . Accessed October 30th 2009.
11) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. ‘‘AHRQ Quality Indicators. Prevention Safety Indicators: Technical Specifications, Version 3.2’’: http://www.qualityindicators.ahrq.gov/modules/PSI_TechSpec.aspx
12) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." Annals of Internal Medicine 158.5_Part_2 (2013): 369-75. Web.
13) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." Annals of Internal Medicine 158.5_Part_1 (2013): 350-52. Web.
14) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." Annals of Internal Medicine 158.5_Part_2 (2013): 365-69. Web.
15) Umscheid CA, Mitchell MD, Agarwal R, et al, Mortality from Reasonably-Preventable Hospital-Acquired Infections, (Philadelphia Penn Center for Evidence-based Practice Advisory, 2008). http://www.shea-online.org/View/smid/428/ArticleID/178.aspx.
16) Ranji SR, Shetty K, Posley KA, et. Al, Prevention of Healthcare-Associated Infections. Vol 6 of Shojania KG, McDonalds KM, Wachter RM, Owens DK, editors. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Technical Review 9 (Prepared by the Stanford University – UCSF Evidence-based Practice Center under Contract No. 290-02-0017). AHRQ Publication No. 04(07)-0051-6. Rockville, MD: Agency for Health Research and Quality. January 2007.
17) Cuddigan J, Berlowitz DR, Ayello EA. “Pressure Ulcers in America: Prevalence, Incidence, and Implications for the Future: An Executive Summary of the National Pressure Ulcer Advisory Panel Monograph.” Advances in Skin and Wound Care 14.4 (2001): 208-215.
18) Leizorowicz A, Cohen AT, Turpie AF, et al. “Randomized placebo-controlled trial of dalteparin for the prevention of venous thromboembolism in acutely ill medical patients.” Circulation 110 (2004): 874-9.
19) Hull, R. D., S. M. Schellong, V. F. Tapson, M. Monreal, M. M. Samama, A. G. G. Turpie, and R. D. Yusen. 2007. ‘‘Extended-Duration Venous Thromboembolism Prophylaxis in Acutely Ill Medical Patients with Recent Reduced Mobility: The EXCLAIM Study.’’ 2007 Congress of the International Society on Thrombosis and Hemostasis; July 7–13, 2007; Geneva, Switzerland. Late-breaking clinical trials, abstract O-S-001.
20) Singri, N., S. N. Ahya, and M. L. Levin. 2003. ‘‘Acute Renal Failure.’’ Journal of American Medical Association 6 (289): 747–51.
21) Landro, L. 2009. ‘‘An Affordable Fix for Modernizing Medical Records’’ Available at http://www.wsj.com/articles/SB124104350516570503
22) Dharmarajan, Kumar, et al. "Diagnoses and Timing of 30-Day Readmissions After Hospitalization for Heart Failure, Acute Myocardial Infarction, or Pneumonia." JAMA 309.4 (2013): 355. Web.
23) Jiang HJ, Russo CA, and Barrett ML. Nationwide Frequency and Costs of Potentially Preventable Hospitalization, 2006. AHRQ-HCUP Statistical Brief # 72. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.jsp.
24) MedPac. Report to the Congress: Promoting Greater Efficiency in Medicare. Rep. Washington DC: Medicare Payment Advisory Commission, 2007. Web.
25) S.F. Jencks, M.V. Williams, and E.A. Coleman: Rehospitalizations among Patients in the Medicare Fee-for-Service Program. New England Journal of Medicine 2009; 360: 1418-1428.
26) Jack BW, Chetty VK, Anthony D et al. “A Reengineered Hospital Discharge Program to decrease Rehospitalization: A randomized trial.” Ann Int Med 150 (2009):178-187.
27) MedPac 2013. Refining the Hospital Readmissions Reduction Program. Rep. no. 4. Vol. 4. Washington DC: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html
28) Lewis G, Kirkham H, and Vaithianathan V. “How Health Systems Could Avert “Triple Fail” Events that are Harmful, are Costly, and result in Poor Patient Satisfaction.” Health Affairs 32.4 (2013): 669-676.
29) de Brantes F. and A. Rastogi, "Evidence-Informed Case Rates: Paying for Safer, More Reliable Care," The Commonwealth Fund 40, publ. 1146 (2008): 1-14.
30) de Brantes, François M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. “Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model.” NEJM (2009): 361:1033 (Perspective)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Pneumonia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=PNE&submit=Submit

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** nqf_pne_all_codes_risk_adjustment_12_14_15.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Measure specifications have been updated since the last endorsement in the following ways:

1. The code tables have been revised to make them more user-friendly and readable. Earlier we had referenced the AHRQ-CCS categories that mapped to the PAC definitions. Now we have displayed the codes as either I-9 or I-10 codes so it is easier for users to use and implement the measure in their own programs.
2. Only codes relevant to Community Acquired Pneumonia have been retained as potential trigger codes. Codes suggestive of a healthcare Acquired Pneumonia such as for ventilator acquired pneumonia or fungal pneumonia and other opportunistic lung infections have been removed.
3. All codes have been updated to 2015 (current codes) and ICD-10 code conversions are included.
4. We no longer define PACs with procedure codes. PAC definitions are based on diagnosis codes and these drive the services for care of the complication. For example, if there is an in-patient infectious disease consultation service for sepsis, the diagnosis code of sepsis on the claim is the tag that alerts the user that there is a complication.
5. Instead of three types of PACs, we now define PACs as one of two types - Type 1 PACs are directly related to the index condition and type 2 PACs are related to patient safety failures. PACs related to comorbidities have been practically eliminated unless they cause patient safety issues, in which case they are listed with type 2 PACs.
6. Our service assignment logic has been modified. All services that are relevant to an episode are multi-assigned to all relevant open episodes. So if a patient had both an open pneumonia episode and an open diabetes episode concurrently, the services relevant to both (such as office visits) will be assigned to both episodes, thereby preventing the possibility of under-counting

services in each episode.
7. We have expanded our databases to include the Medicaid population.
8. Reference to literature and publications have been updated to reflect current knowledge and thinking.
9. Community acquired Pneumonia is now often being managed in an outpatient setting. So we have expanded our measure to include pneumonia cases that were managed completely in an outpatient or office setting, in addition to those that were managed in the inpatient setting. The level of testing has therefore been expanded to include individual physician level testing, in addition to the hospital level testing.

Our team, within HCI3, has been working with various pilot sites across the country to use the PAC (potentially avoidable complications) measures for reporting outcomes at the population level. PAC measures have been overwhelmingly adopted as category 1 quality measures for New York State DSRIP (Delivery System Reform Incentive Payment) project for Medicaid Redesign.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Outcome: Number of patients with pneumonia who had one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The episode starts when the trigger criteria for Pneumonia are fulfilled. The episode time window looks back 7 days from the first trigger claim and continues forward for one month after the trigger date, or for 30-days after discharge if the patient is hospitalized for pneumonia, in order to aggregate relevant claims for the measure.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients with a pneumonia episode that have a potentially avoidable complication (PACs), during the episode time window. The enclosed excel workbook entitled NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls gives the detailed codes for PACs in the tab entitled PACS I-9 & I-10.

Patients are identified as having a PACs if:
a. The index stay for pneumonia has a PAC diagnosis code in any position except in the PRIMARY (principal) position
b. They have a PAC diagnosis code in any position on any relevant claim (outpatient facility, professional, ancillary etc.) during the pneumonia episode time window
c. Any readmission to an acute care facility that is relevant to pneumonia, within the 30-day time window
d. Any admission to a post-acute care facility that is relevant to pneumonia and has a PAC code in any position on the claim

We define PACs as one of two types:
(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a type 1 PAC if they develop one or more complication directly related to pneumonia or its management. Examples of these PACs are respiratory insufficiency, other lung complications, fluid electrolyte acid base problems, sepsis, respiratory failure etc.
(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are considered to have a type 2 PAC, if they develop any of the complications related to patient safety failures such as for phlebitis, deep vein thrombosis, pressure sores or for any of the CMS-defined hospital acquired conditions (HACs).

PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PAC in any of the above settings, they get counted as a "yes" or a 1.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients aged 18 years and above who have a pneumonia episode and are followed for at least one-month.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Please refer to the enclosed excel workbook entitled
NQF_PNE_all_codes_risk_adjustment_12_14_15.xls
The target population is identified based on patients with claims that have a Pneumonia diagnosis codes as defined in the TRIGGERS tab (Triggers I-9 or Triggers I-10) of the enclosed workbook. In addition, they have to meet one of the following trigger criteria:
1. Have a hospitalization with a trigger code in the principal position of an inpatient stay claim
2. Have an outpatient facility visit such as an emergency department visit with one of the trigger codes in any position
3. Have a physician visit with a pneumonia code in any position AND a confirming claim between 7 days and 30 days of the first visit that could be any of the three above (an IP stay claim with a pneumonia code in the principal position, an outpatient facility visit claim or another professional visit claim with the pneumonia diagnosis in any position)

Inclusion criteria: Patients identified to have Pneumonia based on the trigger criteria above are retained in the measure if they meet the following inclusion criteria:
1. The patient has continuous enrollment for the entire time window with no enrollment gaps with the entity providing the data (so we can ensure that the database has captured all the claims for the patient in the time window).
2. The patient has a complete episode time window in the claims data – so the end date of the episode should not be past the database claims end date.
3. Patient is at least 18 years of age

Once the episode is triggered all relevant claims within the episode time window are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Any of these relevant claims serve to identify the presence of a PAC.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The target population captures adult patients (18+) in the dataset, who have a complete episode of community-acquired pneumonia, with no enrollment gaps, and no outlier costs. Patients who do not meet these criteria are excluded from the target population.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Please refer to the tab called "Decision Tree" in the enclosed excel workbook NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls

Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to pneumonia care.

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with the entity providing the data (this helps determine if the database has captured all the claims for the patient in the time window). If a patient has an enrollment gap for any time period during the episode time window, it is considered as an enrollment gap, and they are excluded from the measure.
d. If the pneumonia episode time window extends outside the dataset time period (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events and eliminates random noise into the analysis from inappropriate codes or services. It is also another way to ensure that episodes included in the measure are complete and representative of the measure.

2. "Claims" are excluded from the pneumonia measure if they are considered not relevant to pneumonia care.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

None

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Conceptual Model:
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs may be due to factors that could be controlled by all providers that are managing or co-managing the patient.

Statistical Method:
We use logistic regression to model the probability of at least one PAC occurring during the episode. For each patient the "predicted" coefficients from the risk adjustment model are summed to give the predicted probabilities of the occurrence of a PAC.

A number of patient-related "risk factors" or covariates are included in the model: This list was selected based on input from various clinical experts in clinical working groups. Risk Factors used in the models were:
Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.
Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a potentially avoidable complication (PAC). The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15.xls.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to manage (e.g., morbid obesity) or severity of the illness itself (e.g., viral, gram negative, or MRSA pneumonia). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15.xls.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls). The risk factors along with their codes are listed in the tabs called "All Risk Factors I-9" and "All Risk Factors I-10" and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases
RF0111	Cerebrovascular Disease
RF0115	after care, rehabilitation
RF0201	visual loss, blindness, retinal tear, detachment

RF0301 ENT, Upper Respiratory Problems
RF0401 Respiratory Failure, O2, ventilator dependence
RF0402 Advanced COPD, Asthma
RF0403 Empyema, bronchiectasis, Pneumonias
RF0404 Aspiration Pneumonia, Laryngeal Problems
RF0406 TB, Pneumoconiosis, Aspergillosis
RF0407 Tobacco use, Lung disease due to External Fumes
RF0408 Other Lung Disease
RF0501 Previous Shock, Syncope, Vent Fibrillation
RF0503 Advanced CHF
RF0504 Cardiomyopathy, valve disorders
RF0505 Cardiac Arrhythmias, Heart Block
RF0506 Pacemaker, AICD
RF0507 Endocarditis, Other post surgical cardiac problems
RF0508 Other Cardiovascular Disease
RF0511 DVT, Pulm Embolism, Pulm Heart Disease
RF0512 Unstable Angina
RF0513 Hypotension, chronic, orthostatic
RF0514 Hyperlipidemia
RF0515 Intraaortic Balloon Pump
RF0516 ventricular assist device, ecmo, prolonged bypass
RF0517 Previous electrophysiology studies, cryoablation
RF0518 Recent AMI
RF0519 Previous PCI
RF0520 Previous CABG
RF0521 Previous Heart & Valve Surgery
RF0522 Previous aortic reconstruction
RF0523 Previos carotid endarterectomy
RF0524 Aortic and peripheral vascular disease
RF0525 Advanced Aortic and Vascular Disease
RF0601 GI Bleed
RF0602 Intestinal Obstruction/Perforation
RF0603 Acute Gastritis, Duodenitis
RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas
RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis

RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control
RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders
RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix
RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers
RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers

RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs
RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns
RF2301 Major Organ Transplant Status
RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma
RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for pneumonia

STDX04138	Viral Pneumonia
STDX04171	Influenza w pneumonia
STDX04172	Gram Negative Pneumonia
STDX04173	MRSA Pneumonia
STDX04174	Other Staph Pneumonia
STDX1019	Morbid Obesity (concurrent)
STDX10107	Obesity (concurrent)
STDX1007	Overweight (concurrent)
STDX10108	Sleep Apnea (concurrent)

As you may notice some of the covariates (risk factors) such as obesity are collected from both historical claims as well as from the index stay and look-back period of the episode.

The prevalence of the risk factors in our analysis dataset are listed in the enclosed workbook entitled NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls – see tab “Risk Factor Prevalence”.

The regression model with its coefficients are given in the same workbook in the tab ‘Risk Model’.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Available in attached Excel or csv file at S.2b

S.16. Type of score:
Rate/proportion
If other:

S.17. **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

S.18. **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.)*
Please refer to the enclosed excel workbook entitled (NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls).

Assembling the Denominator:

Using administrative claims database, patients with pneumonia are identified as those who fulfilled the trigger criteria for pneumonia. Pneumonia patients should have claims that have a Pneumonia diagnosis codes as defined in the TRIGGERS tab (Triggers I-9 or Triggers I-10) of the enclosed workbook. In addition, they have to meet one of the following trigger criteria:

1. Have a hospitalization with a trigger code in the principal position of an inpatient stay claim
2. Have an outpatient facility visit such as an emergency department visit with one of the trigger codes in any position
3. Have a physician visit with a pneumonia code in any position AND a confirming claim between 7 days and 30 days of the first visit that could be any of the three above (an IP stay claim with a pneumonia code in the principal position, an outpatient facility visit claim or another professional visit claim with the pneumonia diagnosis in any position)

Patients are retained if they are 18 years of age or more, do not have a missing gender, have continuous enrollment for the entire episode time window, and their entire time window is covered in the claims dataset.

Once the episode is triggered all relevant claims within the episode time window are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Any of these relevant claims serve to identify the presence of a PAC.

Readmissions carrying diagnosis codes relevant to pneumonia, and relevant admissions to post-acute care facilities are also included in the episode. If a patient has more than one concurrent episode open, and the claim is relevant to both episodes, the claim gets multi-assigned to all relevant open episodes preventing undercounting of PACs.

Once all the episodes are assembled, episodes that have outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis. This retains episodes that are more representative of care around pneumonia and excludes episodes that may be incomplete (low outlier costs), or have inappropriate codes or services leading to high outlier costs.

Assembling the Numerator:

For every episode included in the denominator, episodes are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

- Any Index stay that has a PAC diagnosis code in any position except in the PRIMARY (principal) position
- Any readmission to an acute care facility 2 days or later after discharge but within 30-days post-discharge
- Any admission to a post-acute care facility with a PAC code in any position on the claim
- Any other service (professional, outpatient facility, ancillary) with a PAC code in any position on the claim

Relevant claims that do not qualify as a PAC based on the criteria outlined above, are listed as typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of pneumonia patients that have a PAC is simply the ratio of patients with PACs within the pneumonia population, and is called the PAC rate as shown in the equation below:

PAC rate = Patients with pneumonia that have at least one PAC / Total number of pneumonia patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled “Decision Tree” of the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled “PAC overview”, not only do we have the PAC rate for the entire pneumonia population analyzed (54.7%), we can calculate the frequency of PACs occurring in the hospital setting, in the outpatient facility, or in professional claims. These could be further broken down by the PAC type – type 1 being directly related to pneumonia and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement by hospitals and nursing facilities (see tab labeled as “PAC Drill down Graph”). Additionally, readmissions could be analyzed separately. This helps focus strategies in reducing PACs and makes the data immensely actionable.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness using subtypes collected from the trigger claim and / or look-back period. This helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more PACs (=1) or not (=0).

Independent Variables:
A number of patient-related “risk factors” or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient’s lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient’s risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled “All Risk Factors I-9” and “All Risk Factors I-10” for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., viral, gram negative, or MRSA pneumonia). Please see the tab labeled “Subtypes I-9” and “Subtypes I-10” for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15 .xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods:
We use logistic regression to model the probability of at least one PAC occurring during the episode. For each patient the “predicted” coefficients from the risk adjustment model are summed to give the “patient-level” predicted probabilities of the occurrence of a PAC. Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC).

To prevent unstable coefficients, comorbidities and subtypes are included in the models as covariates only if they are present in at least 10 episodes. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be

individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the attribution rules. For community acquired pneumonia episodes, where the index claim is in the hospital setting, the episode is attributed to the facility where the index hospitalization occurred. In a second attribution exercise, all community acquired pneumonia episodes are attributed to the physician who has the maximum number of E&M claims during the episode time window.

To directly compare PAC rates across facilities or physicians while also appropriately accounting for differences in patient severity, we calculate a risk-standardized PAC rate (RSPR) for each provider. This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.).

1. For each provider, the actual number of PAC occurrence is summed across all attributed pneumonia patients, to give the observed PAC rates for the provider.
2. Similarly, patient-level probability estimates are summed across all attributed patients to give expected PAC rates for the provider.
3. The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.
4. To facilitate accurate comparisons of rates across providers, the O/E ratio is multiplied by the overall expected PAC rate across all facilities or physicians, to obtain the risk-standardized PAC rate (RSPR) for the facility or physician.

The formula for this calculation is as follows:

$$RSPR\_j=\{(SUM\ Observed\_ij\ )/(SUM\ Prob\_ij\ )\} \times \{(SUM\ Prob\_i) / (\#\ of\ episodes)\}$$

Where an individual i is attributed to the unit of attribution j (e.g., facility, physician, etc.)

The risk-standardized PAC rate (RSPR) therefore adjusts the provider's observed PAC rate, by the severity of the panel of their patients. It represents what a provider's PAC rate would be if their patient population was reflective of the overall population, leveling the playing field, and allowing for meaningful comparisons across all providers adjusted similarly.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not Applicable

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If data is missing, the case is deleted from both the numerator and denominator

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has 3,258,706 covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization.

The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual, Facility, Population : Regional

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Ambulatory Care : Urgent Care, Hospital/Acute Care Facility, Other
If other: Across the care continuum

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0708_PNE_Testing_Reliability_Validity_HCI3_12_14_2015_updated.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0708
Measure Title: Proportion of Patients with Pneumonia that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 02/14/16
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions
- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more***

than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.

- **For <u>all</u> measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specifications</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*including questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

<u>Note</u>: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record

☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing?
April 1, 2012 – December 17, 2014

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☒ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Pop: Regional	☒ other: Pop: Regional

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)
We included 82 facilities and 170 physicians for testing and analysis. Pneumonia episodes were attributed to the treating facilities in cases where the patients were hospitalized, and in a second attribution exercise to physicians who were primarily responsible for managing the pneumonia (those with the maximum number of E&M services). Because facilities or physicians with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
There were a total of 13,228 episodes of pneumonia included in the testing and analysis. Patients in these episodes were, on average, 57.1 years of age (range 18-64) and 45.7% were female. We did not have race information on these patients. All patients for this analysis fulfilled the trigger criteria as defined in the specifications (Section S.9).

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

For the reliability analysis, we restricted the data to facilities and physicians with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the facility or physician to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
None of the SDS variables were available in the dataset analyzed.

2a2. RELIABILITY TESTING
***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

Reference: Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

The tables below provide summaries of the reliability score for different minimum sample size thresholds. For complete results, refer to the workbook entitled, NQF_PNE_all_codes_risk_adjustment_12_14_15, under the "Facility RSPR & Reliability" and "Physician RSPR & Reliability" tabs to see provider-specific results.

Facilities

Reliability Scores	**Minimum # Episodes Per Facility**	
	>=10	**>=125**
# of Providers (%)	82 (100)	15 (18)
Median (IQR)	0.56 (0.34,0.69)	0.78 (0.73,0.82)
Range	0.16-1.00	0.71-0.84

Physicians

Reliability Scores	**Minimum # Episodes Per Provider**

	>=10
# of Providers (%)	170 (100)
Median (IQR)	0.85 (0.78,0.92)
Range	0.66-1.00

Additionally, the table below compares the minimum sample sizes required for the various PAC measures to achieve an "absolute" reliability score of > 0.7. The table also gives the minimum sample size required to achieve a "median" reliability score > 0.7 in the dataset studied, for both facilities and physicians. On an average, higher sample sizes are required for facilities because of the lower variability in rates of complications across facilities (please see graphs in the attachment called PACs and Severity Adjustment Fact Sheet.docx).

Unit	Sample Size for "Absolute" Reliability >0.7	Sample Size for "Median" Reliability >0.7
Facility	125	80
Physician	11	10

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within a provider's panel) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean. Values above 0.9 are considered sufficient to see differences between pairs of physicians or pairs of facilities (see Adams, 2009 cited above).

Facilities
Among facilities with at least 10 episodes, reliability scores were generally low, with only 25% of scores above the minimum accepted level of 0.7. However, for facilities with a pneumonia volume of at least 80 patients, the median reliability score was 0.7, and for facilities where the volume of pneumonia patients exceeded 125 attributed patients, the reliability scores were consistently above 0.70, demonstrating that for facilities with a high number of episodes the pneumonia PAC measure sufficiently differentiates facility performance (See "Facility RSPR & Reliability" tab of the attached workbook).

Physicians
Among physicians with at least 10 episodes, more than three-quarters had reliability scores of 0.78 or higher, well above 0.70. Moreover, just 10% of physicians in the sample had scores below this threshold. This strongly indicates that the measure more than adequately differentiates performance, using physician PAC rates, across a wide range of sample sizes. (See "Physician RSPR &Reliability" tab of the attached workbook).

2b2. VALIDITY TESTING
2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)
☐ **Performance measure score**
☐ **Empirical validity testing**
☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or

resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
We illustrate that our measure has face validity in several ways.

First, development of the PAC definitions involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help defined and review all of the diagnosis codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship

between low PACs – fewer "defects" – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Given the significant clinical input that went into developing the measures, the widespread use and acceptance the measures have gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-14], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Dec 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

14. Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

There were no real exclusions. The target population is designed to capture adult patients (18+) in the dataset, who have a complete episode of community-acquired pneumonia, with no enrollment gaps, and no outlier costs. Patients who do not meet these criteria are excluded from the target population. Therefore, denominator exclusions included exclusions of "patients" as well as "claims" not relevant to pneumonia care.
1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with the entity providing the data (this helps determine if the database has captured all the claims for the patient in the time window). If a patient has an enrollment gap for any time period during the episode time window, it is considered as an enrollment gap, and they are excluded from the measure.
d. If the pneumonia episode time window extends outside the dataset time period (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events and eliminates random noise into the analysis from inappropriate codes or services. It is also another way to ensure that episodes included in the measure are complete and representative of the measure.
2. "Claims" are excluded from the pneumonia measure if they are considered not relevant to pneumonia care.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
After selecting patients with a valid pneumonia trigger and a complete pneumonia episode, the pneumonia population included in the analysis consisted of 13,228 episodes. As mentioned above, no real exclusions were done. As such, no formal exclusion testing was done.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
No formal analysis was done on the impact of exclusions on performance scores.
Descriptive Explanation:
Exclusions of patients were for comparative purposes or for medical reasons.
(a) Comparative Purposes: We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.
(b) Medical Reasons: Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13 What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model** with 170 potential risk factors and 12 episode specific subtypes
☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.1.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not Applicable

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.
A number of patient-related "risk factors" or covariates are included in the model: This list was selected based on input from various clinical experts in clinical working groups.

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a potentially avoidable complication (PAC). The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15.xls.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to manage (e.g., morbid obesity) or severity of the illness itself (e.g., viral, gram negative, or MRSA pneumonia). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15.xls.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)
All Risk Factors with their coefficients are detailed in the enclosed workbook called NQF_PNE_all_codes_risk_adjustment_12_14_15.xls – Please reference the tabs titled Risk Factor Prevalence and Risk Model.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk(*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Conceptual Methods:

Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, such as patient demographics, comorbidities, severity of illness, complexity of the procedure etc., the remaining variance in PAC (potentially avoidable complication) rates are probably due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window. Because our measure of PACs is based on the occurrence of one or more PACs during the episode, the severity-adjustment models are intended to predict the probability of having at least one PAC during an episode given the patient's combination of risk factors and comorbidities. To

avoid creating perverse incentives, the models consider only those comorbidities and indicators of episode severity that are present at the start of the episode. None are identified during the episode period.

Clinical Methods:

The list of risk factors and subtypes was selected based on input from clinical experts in clinical working groups.

Statistical Methods:

The *unit of analysis* is the individual episode.

The *dependent variable* is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables include patient demographic factors, historic risk factors as indicators of patient comorbidities and subtype markers as indicators of patient's severity of illness.

Statistical Modeling: We use logistic regression to model the probability of at least one PAC occurring during the episode. For each patient the "predicted" coefficients from the risk adjustment model are summed to give the predicted probabilities of the occurrence of a PAC.

To prevent unstable coefficients, comorbidities and subtypes are included in the models as covariates only if they are present in at least 10 episodes. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the models be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of an episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode cost models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tab titled Risk Model in the NQF_PNE_all_codes_risk_adjustment_12_14_15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
Not Applicable since our analysis did not include SDS variables.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Model Development Approach
We used logistic regression to model the probability of having at least one PAC during the episode. PACs could occur anytime during the episode, either during the IP stay, in the post-discharge phase, as a readmission, or

identified with a PAC code on any relevant claim during the episode time window. The model included all covariates that were either collected as risk factors from historical claims before the start of an episode, or were present as subtype variables on the trigger claim or during the look-back period. No further model building was conducted after the initial model was run.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample). The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

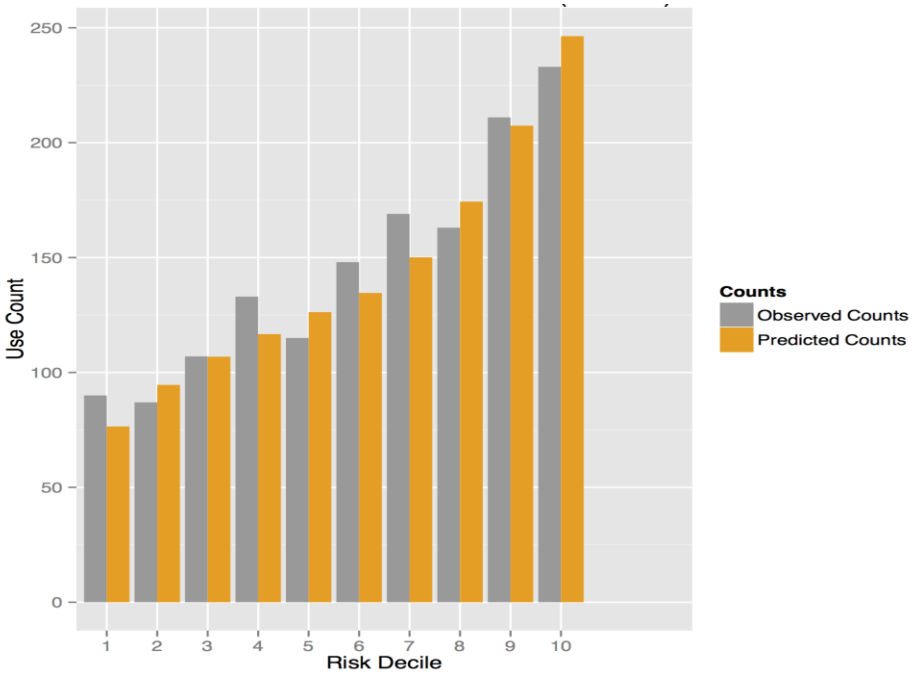
Sample	Accuracy (%)*	AUC (C-Statistic)
Development	64.7%	0.700
Validation	65.8%	0.720

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC).

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Development	30.4972	8	<0.0001
Validation	8.1555	8	0.4184

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could perhaps be controlled by all providers that are managing or co-managing the patient, during the entire episode time window. The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1].

The results above indicate that the C-statistics for the risk model on the development and validation samples (0.700 and 0.720, respectively) are at the level at which the model is considered to have good discriminatory power. Moreover, the accuracy values show that the testing model correctly predicts whether an episode had or did not have a PAC around 65% of the time, well above what would be expected if the predictions were made at random (i.e., 50%). Although the H-L test was significant for the development sample, this test is generally known to be sensitive to the number of groupings used and sample sizes. Additionally, the risk decile plot shows that the model predicts occurrence of PACs quite similar to the number of observed PACs, across most of the deciles.

Overall, the results demonstrate that the model has good predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across facilities or physicians while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate (RSPR) for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. Using the risk-adjustment model, we calculated the expected number of episodes with PACs for each provider's panel of attributed patients, given their patients risk profile. We then calculated the ratio of observed attributed episodes with PACs, to the expected PACs, for each facility or physician. This number yielded whether the facility or physician had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all facilities or physicians, to obtain the risk-standardized PAC rate (RSPR) for the facility or physician. This measure therefore adjusts the provider's observed PAC rate, by the severity of the panel of their patients. It represents what a provider's PAC rate would be if their patient population was reflective of the overall population, leveling the playing field, and allowing for meaningful comparisons across all providers adjusted similarly.

Because providers with small volumes may provide unreliable estimates, we excluded any providers with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across facilities treating pneumonia (see "Facility RSPR & Reliability" tab in the enclosed workbook), and in a separate attribution exercise we compared risk-adjusted PAC rates across physicians treating pneumonia (see "Physician RSPR & Reliability tab" in the enclosed workbook). We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

References:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Please refer to the NQF_PNE_all_codes_risk_adjustment_12_14_15.xls workbook under the "Facility RSPR & Reliability" tab and "Physician RSPR & Reliability" tab to see specific results for each facility and physician respectively.

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

Facilities:

PAC Rates	Minimum # Episodes Per Facility	
	>=10	>=125 (Reliability >0.7)

Unadjusted		
Median (IQR)	63% (58%, 69%)	63% (58%, 68%)
Range	27%-100%	48%-77%
Adjusted (RSPR)*		
Median (IQR)	63% (57%,69%)	63% (57%, 67%)
Range	30%-91%	50%-78%

Physicians:

PAC Rates	Minimum # Episodes Per Physician >=10 (Reliability >0.7)
Unadjusted	
Median (IQR)	60% (43%, 79%)
Range	0%-100%
Adjusted (RSPR)*	
Median (IQR)	58% (44%, 70%)
Range	0%-100%

*RSPR = Risk Standardized PAC Rate

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The variation in risk-standardized PAC rates suggests there the measure identifies meaningful differences in performance among providers that manage pneumonia at both the level of facilities and physicians.

However, it is important to report risk-standardized PAC rates only for those providers that meet a minimum sample size with reliability scores of >0.7 for that dataset.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not performed

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

No formal analysis was done on the impact of exclusions on performance scores.
Descriptive Explanation:
Exclusions of patients were for comparative purposes or for medical reasons.
(a) Comparative Purposes: We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.
(b) Medical Reasons: Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
Not applicable

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
Not applicable

3. Feasibility
Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations that list the data elements required to ensure measure validity; the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure specifications, and contain the coding information required, the analysis of the measure and its results are highly reliable.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the

time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/ Horizon Blue Cross Blue Shield of New Jersey http://www.horizonblue.com/ Pennsylvania Employee Benefits Trust Fund https://www.pebtf.org/ Quality Improvement (Internal to the specific organization) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-9]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform
-Maine Health Management Coalition
-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto. FairHealth has joined others to use our analytics to create PAC rates for various consumer transparency efforts they are engaged in.

More recently, the PAC measures have been overwhelmingly adopted as category 1 quality measures for New York State DSRIP

(Delivery System Reform Incentive Payment) project for Medicaid Redesign and are scheduled for pilot site implementations in 2016.

http://www.health.ny.gov/health_care/medicaid/redesign/dsrip/vbp_reform.htm
http://www.health.ny.gov/health_care/medicaid/redesign/dsrip/docs/dsrip_vbp_webinar_slides.pdf

Below are some references that highlight our work with Potentially Avoidable Complications (PACs):
1.Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)
2.Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3.de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4.de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5.Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6.Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7.François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8.de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.
9. de Brantes F and Rastogi A: Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. Commonwealth Fund; 2008 June 17. Available from: http://www.hci3.org/content/evidence-informed-case-rates-paying-safer-more-reliable-care

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,
CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and are working with them to augment provider transparency by using PAC measures,
MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,
Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public dissemination.

In Dec 2014, the measure was conditionally approved by MAP (Measure Applications Partnership), for use in Medicare's Inpatient Quality Reporting program, and continues to be pushed by organizations like the Consumer-Purchaser Alliance for that purpose.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance

results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider network, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually

both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
YesYes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0094 : Assessment of Oxygen Saturation for Community-Acquired Bacterial Pneumonia
0095 : Assessment Mental Status for Community-Acquired Bacterial Pneumonia
0096 : Community-Acquired Bacterial Pneumonia (CAP): Empiric Antibiotic
0141 : Patient Fall Rate
0147 : Initial antibiotic selection for community-acquired pneumonia (CAP) in immunocompetent patients
0148 : Blood cultures performed in the emergency department prior to initial antibiotic received in hospital
0151 : Initial antibiotic received within 6 hours of hospital arrival
0202 : Falls with injury
0232 : Vital Signs for Community-Acquired Bacterial Pneumonia
0233 : Assessment of Oxygen Saturation for Community-Acquired Bacterial Pneumonia
0234 : Assessment of Mental Status for Community Acquired Bacterial Pneumonia
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization
0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1611 : ETG Based PNEUMONIA cost of care measure
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
2579 : Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
-0531 Patient Safety for Selected Indicators (Composite Measure, Endorsed)(AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Denominator Harmonization: Several of the measures listed in the prior section are harmonized to the extent possible for denominator definitions with the submitted measure. In particular process measures related to community-acquired pneumonia (CAP) 0096, 0151, 0147, 0148 have defined CAP target population that matches closely to our submitted measure. Numerator Harmonization: Regarding numerator harmonization, several of the measures are subsets of our measure. In particular 0450, 0337, 0141, and 0202 list adverse events that have been synchronized with those definitions within the PAC measure. In addition, 0705, 0709 have numerator definitions harmonized completely for the definitions of PACs. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measures. While the submitted PAC measures include readmissions that occur within 30 days of discharge, the readmissions, by definition, are related to pneumonia and not due to any cause. While 30-day all-cause readmissions might make sense in a Medicare population, it is not self-evident that they do for commercial or Medicaid populations. However, that said, our data suggest that there are, in fact, very few readmissions within 30 days post discharge that aren't relevant to the index hospitalization. It is worth noting that there is some mounting controversy about the 30 day all cause readmission measures and some data suggest that these measures might have simply pushed out certain readmissions to 31 or more days post discharge. Irrespective of these points, PACs include readmissions and are designed to enable accountability at the

locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events that can occur during the stay, as well as adverse events, including readmissions, that can occur post-discharge. As a result, they provide facilities and physicians with an overall measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.Denominator Harmonization: Several of the measures listed in the prior section are harmonized to the extent possible for denominator definitions with the submitted measure. In particular process measures related to community-acquired pneumonia (CAP) 0096, 0151, 0147, 0148 have defined CAP target population that matches closely to our submitted measure. Numerator Harmonization: Regarding numerator harmonization, several of the measures are subsets of our measure. In particular 0450, 0337, 0141, and 0202 list adverse events that have been synchronized with those definitions within the PAC measure. In addition, 0705, 0709 have numerator definitions harmonized completely for the definitions of PACs. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measures. While the submitted PAC measures include readmissions that occur within 30 days of discharge, the readmissions, by definition, are related to pneumonia and not due to any cause. While 30-day all-cause readmissions might make sense in a Medicare population, it is not self-evident that they do for commercial or Medicaid populations. However, that said, our data suggest that there are, in fact, very few readmissions within 30 days post discharge that aren't relevant to the index hospitalization. It is worth noting that there is some mounting controversy about the 30 day all cause readmission measures and some data suggest that these measures might have simply pushed out certain readmissions to 31 or more days post discharge. Irrespective of these points, PACs include readmissions and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events that can occur during the stay, as well as adverse events, including readmissions, that can occur post-discharge. As a result, they provide facilities and physicians with an overall measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)
Not Applicable

Related Measures: AHRQ-PQIs (PQI 11) Bacterial Pneumonia Admission Rate; CMS-HACs (Hospital Acquired Conditions)Not Applicable

Related Measures: AHRQ-PQIs (PQI 11) Bacterial Pneumonia Admission Rate; CMS-HACs (Hospital Acquired Conditions)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet-635860501570909220.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute
Co.2 **Point of Contact:** Francois, de Brantes, francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute
Co.4 **Point of Contact:** Amita, Rastogi, amita.rastogi@hci3.org, 219-934-9624-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
In 2006 the Prometheus Payment Design Team convened a series of meetings with physicians that had been organized in Clinical Working Groups. These Groups focused on Cancer, Cardiac, Chronic, Orthopedic and Preventive care. The results of this work were summarized in a Commonwealth Fund report published in June 2007 (1) and served as an input to the initial modeling work performed in 2007 and early 2008. The Pneumonia measure was derived from this framework (2).

From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011, 2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications. Subsequently in 2015, HCI3 worked collaboratively with KPMG under the Medicaid DSRIP effort in New York State to get clinical input from Clinical Advisory Groups and Clinical Validation Groups who validated various episode definitions created by HCI3.

For a brief history of PAC measures, please see enclosed document called PAC and Severity Adjustment Fact Sheet.

Some of the clinical experts that have contributed to our work include the following and are also listed at the link below:
http://www.hci3.org/programs-efforts/prometheus-payment/evidence_informed_case_rates/how_is_an_ecr_created/clinical-working-group-contributors

-Dr. John Allen, American Gastroenterology Association (AGA)
-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL
-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA
-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)
-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL

-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

_____

(1) de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model. Commonwealth Fund; 2007 May 20. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278.

(2) de Brantes F and Rastogi A: Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. Commonwealth Fund; 2008 June 17. Available from: http://www.hci3.org/content/evidence-informed-case-rates-paying-safer-more-reliable-care

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2008

Ad.3 **Month and Year of most recent revision:** 12, 2015

Ad.4 **What is your frequency for review/update of this measure?** Yearly

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** Please see enclosed document called PAC and Severity Adjustment Fact Sheet attached to Appendix A.1

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1799
De.2. **Measure Title:** Medication Management for People with Asthma
Co.1.1. **Measure Steward:** National Committee for Quality Assurance
De.3. **Brief Description of Measure:** The percentage of patients 5-64 years of age during the measurement year who were identified as having persistent asthma and were dispensed appropriate medications that they remained on during the treatment period. Two rates are reported.
1. The percentage of patients who remained on an asthma controller medication for at least 50% of their treatment period.
2. The percentage of patients who remained on an asthma controller medication for at least 75% of their treatment period.
1b.1. **Developer Rationale:** This measure assesses adherence to long-term asthma controller medications in patients with persistent asthma. The improvement in quality envisioned by the use of this measure is increasing adherence to long-term asthma controller medications in patients with persistent asthma. According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered "not well controlled" or "very poorly controlled" as defined by clinical practice guidelines (Stillman 2010). Increasing adherence to asthma controller medications can prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami 2009; (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007).

Akinbami, L.J., J.E. Moorman, P.L. Garbe, E.J. Sondik. 2009. Status of Childhood Asthma in the United States, 1980–2007. Pediatrics 123;S131-45. doi: 10.1542/peds.2008-2233C.

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf (November 19, 2015).

Stillman, L. 2010. Living with Asthma in New England: Results from the 2006 BRFSS and Call-back Survey. A report by the Asthma Regional Council of New England (February). http://www.hria.org/uploads/catalogerfiles/living-with-asthma-in-new-england/HRiA_Living_with_Asthma_BRFSS_2010.pdf (November 19, 2015).

S.4. **Numerator Statement:** Numerator 1 (Medication Adherence 50%): The number of patients who achieved a PDC* of at least 50% for their asthma controller medications during the measurement year. A higher rate is better.

Numerator 2 (Medication Adherence 75%): The number of patients who achieved a PDC* of at least 75% for their asthma controller medications during the measurement year. A higher rate is better.

*PDC is the proportion of days covered by at least one asthma controller medication prescription, divided by the number of days in the treatment period. The treatment period is the period of time beginning on the earliest prescription dispensing date for any asthma controller medication during the measurement year through the last day of the measurement year.
S.7. **Denominator Statement:** All patients 5–64 years of age as of December 31 of the measurement year who have persistent asthma by meeting at least one of the following criteria during both the measurement year and the year prior to the measurement year:
• At least one emergency department visit with asthma as the principal diagnosis
• At least one acute inpatient claim/encounter with asthma as the principal diagnosis
• At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits.

• At least four asthma medication dispensing events

S.10. **Denominator Exclusions:** 1) Exclude patients who had any of the following diagnoses any time during the patient's history through the end of the measurement year (e.g., December 31):
-COPD
-Emphysema
-Obstructive Chronic Bronchitis
-Chronic Respiratory Conditions Due To Fumes/Vapors
-Cystic Fibrosis
-Acute Respiratory Failure

2) Exclude any patients who had no asthma controller medications dispensed during the measurement year.

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy
S.26. **Level of Analysis:** Health Plan, Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2012 **Most Recent Endorsement Date:** Jul 31, 2012

Maintenance of Endorsement Measure -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary or **Summary of prior review in 2012**

The developer provides the following evidence for this measure:

- The Level of Analysis is Health Plan/ Integrated Delivery System
- During the previous review, the Committee agreed the evidence supports the importance of adherence to medication regimens in controlling asthma and that adherence to medication regimens in controlling asthma results in fewer exacerbations, ED visits, and hospitalizations. However, the Committee noted concern regarding a lack of evidence for the 50% and 75% proportion of days covered (PDC) threshold values in relationship to outcomes.
- Original evidence for this process measure is based on Clinical Practice Guideline recommendations (National Heart Lung and Blood Institute/National Asthma Education and Prevention Program 2007 Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma); these guidelines, dated 2007, referenced 1,654 studies to update a previous set of guidelines from 2004. The guidelines referenced 529 studies related to pharmacological therapy for asthma including meta-analyses, systemic reviews of RCTs, case control and cohort studies and non-analytic studies including case reports and case series. Although guideline developers did not provide a breakdown of specific numbers of RCTs, the evidence related to inhaled corticosteroid dosing included 34 RCTs or meta-analyses and systemic reviews of RCTs. The 34 RCTs referenced included thousands of patients studied over long periods of time. The evidence supporting the recommendation of daily long-term asthma

controller medication for patients with persistent asthma is strong, Category A, which includes randomized controlled trials (RCTs), a rich body of data.

- NHLBI Overall Recommendation: Adherence to long-term asthma controller medications that are taken daily on a long-term basis help patients achieve and maintain control of persistent asthma. Increasing adherence to asthma controller medications can prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami 2009; (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007).
- Recommendation 1: Daily long-term control medication is recommended for patients who have persistent asthma. The long-term control medication should be ones with anti-inflammatory effects. Of the available medications, ICSs are the most effective single agent. (Evidence Category A).
- Recommendation 2: When initiating therapy, monitor at 2- to 6- week intervals to ensure that asthma control is achieved (Evidence Category D)
- Recommendation 3: Regular follow-up contacts at 1- to 6- month intervals, depending on the level of control, are recommended to ensure that control is maintained and the appropriate adjustments in therapy are made. (Evidence Category D)
- Recommendation 4: Long-term control medications are taken daily on a long-term basis to achieve and maintain control of persistent asthma. The most effective long-term-control medications are those that attenuate the underlying inflammation that is characteristic of asthma. (Evidence Category A)

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- The developer conducted a literature search for published peer-reviewed journals related to the correlation between asthma controller medication adherence and improved outcomes.
  - Specifically using this HEDIS measures, patients who achieved 75% compliance in 2012 did not have fewer hospitalizations or ED visits in 2013 compared to those who were not 75% compliant. Patients who achieved 50% threshold in 2012 did not have fewer hospitalizations, but did have fewer ED visits in 2013, compared to those who were not 50% compliant. (Yoon 2015)
  - Patients with asthma controller medication adherence rates of 75% or greater of the prescribed dose had a statistically significant reduction in asthma exacerbations compared to patients whose controller adherence was 25% or less. (Williams 2011)
  - Based on a 75% medication possession ratio, more-compliant patients were less likely to experience exacerbation than less-compliant patients. (Stern 2006)
  - In pediatric patients with asthma, there were statistically significant differences in adherence rates for maintaining or not maintaining the asthma control. Optimal asthma control entailed adherence rate higher than 80%. (Lasmar 2009)
  - Adherence to inhaled corticosteroids was significantly and negatively correlated with the number of ED, the number of fills of an oral steroid, and the total days' supply of oral steroid. During the two-year study period, there were 80 asthma-related hospitalizations; an estimated 32 hospitalizations would have occurred were there no gaps in medication use. (Williams 2004)

Exception to evidence: Not applicable

Guidance from the Evidence Algorithm: 1->3->4->5 of the algorithm (eligible for a HIGH rating)

Questions for the Committee:

- *The previous Committee expressed concern about the lack of evidence related to the 50% and 75% thresholds for*

proportion of days covered by medication use (PDC)? The literature review conducted by measure developer yielded one new study since the last NQF review. How should the evidence be viewed in light of the 2015 article?

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer noted the following:

- According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered "not well controlled" or "very poorly controlled" as defined by clinical practice guidelines (Stillman 2010).
- Data were extracted from the NCQA Healthcare Effectiveness Data and Information Set (HEDIS) reflecting the most recent years of measurement for this measure at the health plan level (commercial HMO and PPOs; Medicaid) for both numerator types (50% and 75% medication adherence):
  - The mean results for 50% Medication adherence ranged from 51%-70% among various plans, with little change seen from 2012-2014 within each plan (</=4%).
  - The mean results for 75% medication adherence ranged from 29%-45% among various plans, with little change seen from 2012-2014 within each plan (</=5%).
  - For the Medication Adherence 50% indicator in 2014, there was a 16 percentage point difference between commercial plans in the 10th percentile and commercial plans in the 90th percentile and 26 percentage point difference for Medicaid plans. Similarly in 2014 for the Medication Adherence 75% indicator, there was a 20 percentage point difference between plans in the 10th percentile and plans in the 90th percentile for commercial plans and 26 percentage point difference for Medicaid plans.

Disparities

- NCQA does not currently collect performance data stratified by race, ethnicity, or language. Medicaid vs. commercial performance is one proxy for disparities. The following is 2014 HEDIS data from the most recent years of measurement (2012-2014):

1. Medication Adherence 50%:
   - Mean Commercial Rate= 69% with </=4% change
   - Mean Medicaid Rate= 54% with </=3% change

2. Medication Adherence 75%:
   - Mean Commercial Rate= 45% with </=5% change
   - Mean Medicaid Rate=31% with </=2% change

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Do the data demonstrate a potential for improvement in asthma management and exacerbation rates?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The literature review conducted by measure developer yielded one new study since the last NQF review. Concern will likely again be expressed about the evidence related to 50% and 75% following the Yoon study.

**There is a relationship between the measured outcome and a healthcare action; it applies directly to the outcome. Evidence supports adherence to medication regimens in controlling asthma results in fewer exacerbations, ED visits, and hospitalizations.

**This is a process measure, supported by numerous clinical practice guidelines that indicate a link between appropriate medication management and decreased asthma exacerbation.

1b. Performance Gap

Comments:

**2014 data clearly shows a performance gap which warrants a national performance measure.
**There is a disparity in adherence rates for the Commercial vs. the Medicaid populations.
**Performance data was noted for varying levels of medication adherence. Disparity data was not collected, but HEDIS data served as a proxy, indicating differences between Medicaid and commercial payer rates of exacerbation.

1c. High Priority (previously referred to as High Impact)
Comments:
**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy

Specifications:

- Developer attests: "There are no significant changes to the measure specifications since the last endorsement maintenance completed on January 31, 2012."
- This measure is a process measure and is not risk adjusted.
- The measure specifies two separate numerators (50% and 75%) of medication management (proportion of days covered [PDC])
- The denominator specifies four criteria, at least one of which must be met in the measurement year and the year prior to the measurement year. The criteria appear straightforward.
- ICD-9 and ICD-10 codes have been included in specification details

Question for the Committee:

- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- In the 2012 review, the developer conducted field testing in 9 health plans, including both commercial and Medicaid plans, and with membership ranging from 2,000 to 700,288.
- The previous Steering Committee rated reliability as medium.

Describe any updates to testing:

- Updates have been made to the reliability testing since the last submission. Additional empirical validity testing of the measure score has been conducted with data from health plans that submitted HEDIS data for this measure in 2012 and 2015.
- Developers provided 2015 measure score reliability using data for measurement year 2014 (measurement year 2014 required two years' worth of health plan data from January 1, 2013 through December 31, 2014). The updated testing involved 388 Commercial health plans and 192 Medicaid health plans.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing:

- The developer conducted beta-binomial testing, which it states provides a better fit when estimating the reliability of simple pass/fail rate measures. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). The developer states that, generally, a minimum reliability score of 0.7 indicates sufficient signal strength to discriminate performance between accountable entities.

Results of reliability testing:

- The overall beta-binomial statistics for each indicator for commercial and Medicaid sites follow:
  - Medication Adherence 50%: Commercial = 0.84; Medicaid = 0.93
  - Medication Adherence 75%: Commercial = 0.87; Medicaid = 0.97
  - Per the developer, the 10-90th percentile distribution of health plan level-reliability on the rates in this measure show the vast majority of health plans exceeded the minimally accepted threshold of 0.7, and the majority of plans exceeded 0.8.

Guidance from the Reliability Algorithm: 1→2→3→4→5→6/7 (eligible for HIGH rating)

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**

- During previous review, the Committee questioned whether the measure is consistent with evidence—specifically whether there was sufficient evidence for the 50% and 75% adherence thresholds in the numerator.
  - The developer explained these threshold values were based on expert panel input. Additionally, the developer stated one field test site examined ED visits for the population below and above 50% adherence, finding higher ED use in the <50% population.
  - Two new studies since the previous review were reported by the developer.
- The Committee questioned calculating the "proportion of days covered (PDC)" based on an Index Prescription Start Date [IPSD], the earliest dispensing date.
- The Committee identified a challenge for Medicaid patients meeting the 2-year persistent asthma definition due to transient enrollment.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- NCQA previously tested the measure results for face validity using a panel of experts (Respiratory Advisory Panel-RMP). This measure was deemed valid by the expert panel and approved by NCQA's Committee on Performance Measurement for inclusion in HEDIS.

Describe any updates to validity testing:

- Since the last NQF review, the developer has conducted empirical validity testing at the level of the performance score.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- Construct Validity: The developer examined whether the two indicators were correlated with other similar measures of respiratory care (i.e., other similar HEDIS measures). The Pearson correlation test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.
- Pearson Correlation Coefficients (PCC) were calculated for data from 388 Commercial plans for the 2015 HEDIS measure (use 2014/2013 data).

Validity testing results:

- Empirical Validity Testing (Construct):
  - Per the developer, the results (n=388 commercial plans) indicated the asthma measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized. The developer posits the level of correlations among 2 of the 3 indicators for both Commercial and Medicaid plans ranged from moderate to strong (PCC higher than 0.3).
    - Medication Adherence 50% and Medication Adherence 75%: PCC=0.9
    - Medication Adherence 50% and Asthma Medication Ratio: PCC=0.3
    - Medication Adherence 75% and Asthma Medication Ratio: PCC=0.2

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer notes the following:

- No new exclusion testing was performed. The developer reports the following from initial development in 2010:
  1. To ensure the measure captures people with persistent asthma only and not intermittent or seasonal asthma, members who did not meet one of four criteria (including asthma medication dispensing events and/or encounters with an asthma diagnosis) in both the measurement year and the year prior were excluded
  2. People with a diagnosis for a specific clinical condition (COPD, emphysema, obstructive chronic bronchitis, cystic fibrosis and acute respiratory failure) were excluded.
- Exclusions were tested using data from 7 commercial and 7 Medicaid health plans to determine the impact each clinical condition had on the measure's performance.

- The percent of people excluded from the denominator (i.e., the percent who had at least 1 exclusion condition) and the total percent of people excluded from the denominator for each age cohort is provided.
- Overall, 25% OF Commercial members and 18% of Medicaid members were excluded.
- A higher percentage of people ages 51-64 had at least 1 measure exclusion compared to children or adults age 5-50 years.
- The denominator criteria and the exclusions in this measure are intended to focus the measure on children and adults who have persistent asthma rather than concomitant diagnoses of asthma and COPD or chronic bronchitis. The higher percentage of adults ages 51-64 being excluded from the measure was expected, as they have a higher prevalence of conditions such as COPD or chronic bronchitis.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Question for the Committee:

- *Is the lack of risk adjustment appropriate?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified):*

- The developer calculates meaningful differences using inter-quartile range (IQR) for each indicator
- The developer reports a 7-14% gap in performance between the 25th and 75th percentile across the different age ranges and product lines. For all product lines and age groups the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance is for Medicaid plans on the Medication Adherence 75% indicator for children ages 5-11, which shows a 14 percentage point gap between 25th and 75th percentile plans.

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

The developer states:

- Plans collect this measure using all administrative data sources.
- NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

Guidance from the Validity Algorithm : 1→2→3→6→7→8 (eligible for HIGH rating)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Specifications are "generally" in sync with the evidence. Note however the thresholds for adherence are based upon expert panel and field testing which may note carry the same weight as formal studies/trials. The new studies reported results that intuitively did not line up. For example, Yoon study only demonstrated fewer ED visits at the 50% compliance level and NOT at the 70% compliance level.

**Consistency is noted as "somewhat", presumably because of threshold values being based on expert panel input, the calculation of PDC, and and transient enrollment challenge.

2a2. Reliability Testing

Comments:

**The results demonstrate sufficient validity so that conclusions about quality can be made and used as an indicator of quality.
**Yes validity was tested with adequate scope. Empirical testing was performed using the two indicators with other similar measures (Asthma medication ratio). Quality conclusions can be made from the differences in performance from the 10 to 90th percentile scores.
**Validity testing was performed at measure score level with empirical validity testing. Noted that medication adherence at 75% does not meet the threshold for moderate to strong level of correlation.

2b2. Validity Testing
Comments:
**NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.
**Lack of risk adjustment may affect results/validity. The sample size spans a large age range and co-morbidities may increase risk of asthma exacerbations.
**The developer notes that measure calculations are not biased due to missing data.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**The results demonstrate sufficient reliability so that differences in performance can be identified.
**Reliability testing did appear to be adequate with added beta-binomial testing using HEDIS scores among plans using the 10-90th percentile distribution.
**Testing was updated to examine 288 commercial and 192 Medicaid plans; reliability testing was only done at the Measure score level with no data element examination.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer noted the following:

- All data elements are defined fields in electronic claims and are generated by or collected by healthcare personnel during the provision of care. The data are coded by someone other than the individual obtaining the original information.
- NCQA conducts an independent audit of all HEDIS collection and reporting processes.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**In use and highly feasible.
**Data elements are generated during routine care. Claims data may not capture all prescription fills, especially if medications are obtained outside of the plan benefits.
**All of the data elements are defined in electronic claims; data is coded by independent parties which may lend itself to human error.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure is publically reported nationally and by geographic regions in NCQA's *State of Health Care* annual report.
- The measure is reported in *Consumer Reports* and on the NCQA website and used to calculate health plan ratings.
- The measure is used in scoring for accreditation of Medicare Advantage Health Plans
- This measure is used in *Quality Compass*.
- This measure is used in the Medicaid Child Core Set
- This measure is used in the CMS Health Insurance Market Quality Rating System

Improvement results

The developer provided the following results:

Medication Adherence 50%
Commercial Rate Ages 5-64 (HMO and PPO Combined)
2012 = 66% (mean); range 16-92%
2013 = 70%; 16-93%
2014 = 69%; 46-91%
Medicaid Rate Ages 5-64
2012 = 51%; 23-90%
2013 = 54%; 25-94%
2014 = 54%; 34-87%

Medication Adherence 75%
Commercial Rate Ages 5-64 (HMO and PPO Combined)
2012 = 43% (mean); range 8-82%
2013 = 48%; 7-82%
2014 = 45%; 12-72%;
Medicaid Rate Ages 5-64
2012 = 25%; range 9-77%
2013 = 31%; 9-84%
2014 = 31%; 13-75%

Additional data are provided at 1b2

Unexpected findings (positive or negative) during implementation

- The developer states there were no identified unintended consequences for this measure during testing or since implementation.

Potential harms

- The developer states there were no identified unintended consequences for this measure during testing or since implementation.

Feedback:

Reviewed by MAP:

- Federal Programs: Current Finalized 2012-2013; 2013-2014; 2014-2015: Children's Health Insurance Program Reauthorization Act Quality Reporting
- MAP 2013 Decisions: PQRS-Support
- Child Medicaid Core Set
- MAP Affordability Family Measurement Area
- MAP Affordability Family High-Leverage Opportunity

***Questions for the Committee**:*

- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *Although the measure is in use, is the small (to no) improvement over time indicative of poor usability?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**The benefits of the measure outweigh any potential unintended consequences.
**This measure is being reported by NCQA, and is available publicly. These performance results can be used as quality measures in value based products.
**The measure is currently publicly reported an utilized in accountability programs: NCQA State of Health Care Annual Report, Consumer Reports, NCQA website, Medicare Advantage Health Plan accreditation, Quality Compass, Medicaid Child Core Set, CMS Health Insurance market Quality Rating System. The measure could be utilized to promote medication adherence and monitoring, preventing exacerbations and increased cost of ER utilization.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0047: Asthma: Pharmacologic Therapy for Persistent Asthma
- 1800: Asthma Medication Ratio (AMR)

Harmonization

NCQA Response to Harmonization in current submission: NQF #0047 is not harmonized because it is a physician-level measure that assesses whether a patient was prescribed medication at least once during the measurement year, while NQF #1799 assesses patient adherence to asthma controller medications throughout the measurement year. #0548 is a health plan-level measure that assesses two rates of poor asthma control that indicate over-utilization of rescue medication and need for additional therapeutic intervention; Measure 1799 assesses patient adherence to asthma controller medications during the measurement year. There is no impact on interpretability or added burden of data collection because the focus of each measure is different and the data for each measure are collected from different data sources by different entities.

Pre-meeting public and member comments

- None

Measure Number (*if previously endorsed*)**:** 1799

Measure Title: Medication Management for People with Asthma

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across

Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Adherence to long-term asthma controller medications in patients with persistent asthma

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Members who regularly take their prescribed controller medications experience significantly fewer asthma exacerbations defined as either emergency department (ED) visits with asthma listed as the primary diagnosis. The intent of the measure is to have members be compliant and become an active participant in their own chronic disease management thereby minimizing the number of preventable asthma exacerbations.

Diagram:

Provider identifies patients with persistent asthma >>> Provider dispenses asthma controller medication to patients to be used on a daily basis to manage and control asthma symptoms >>> Provider monitors patients for control, educates patients on the importance of asthma controller medication adherence and assesses patient

adherence to medication >>> Prevention and control of asthma symptoms, improvement in quality of life, and reduction in the frequency and severity of asthma exacerbations (desired outcome).

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051.
http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma.

PERSISTENT ASTHMA

The Expert Panel recommends the following therapy for persistent asthma:

- Daily long-term control medication is recommended for patients who have persistent asthma. The long-term control medication should be ones with anti-inflammatory effects. Of the available medications, ICSs are the most effective single agents (Evidence Category A, page 216).
- Quick-relief medication must be available to all patients who have persistent asthma. SABA should be taken as needed to relieve symptoms. The intensity of treatment will depend on the severity of the exacerbation. Increasing use of SABA or use more than 2 days a week for symptom control (not for preventing exercise-induced asthma) indicates the need to step up therapy. The Expert Panel does not recommend regularly scheduled, daily, long-term use of SABA (Evidence Category A, page 236).

Monitoring and follow-up is essential (Evidence Category B, page 277).

- When initiating therapy, monitor at 2- to 6-week intervals to ensure that asthma control is achieved (Evidence Category D).

- Regular follow-up contacts at 1- to 6-month intervals, depending on level of control, are recommended to ensure that control is maintained and the appropriate adjustments in therapy are made: step up if necessary or step down if possible. Consider 3-month intervals if a step down in therapy is anticipated (Evidence Category D).

LONG-TERM CONTROL MEDICATIONS

The Expert Panel recommends that long-term control medications (including ICSs, inhaled long-acting bronchodilators, leukotriene modifiers, cromolyn, theophylline, and immunomodulators) be taken daily on a long-term basis to achieve and maintain control of persistent asthma. The most effective long-term-control medications are those that attenuate the underlying inflammation that is characteristic of asthma (Evidence Category A, page 216).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

- When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel.
- Category A: Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.
- Category B: RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.
- Category D: Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

- When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong.
- Category C: Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051.
http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

This measure assesses adherence to long-term asthma controller medications in patients with persistent asthma. The measure is based on guidelines and evidence that long-term asthma controller medications that are taken daily on a long-term basis help patients achieve and maintain control of persistent asthma.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

- When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel.
- Category A: Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.
- Category B: RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.
- Category D: Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

- When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong.
- Category C: Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1997-2006

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The Guidelines for the Diagnosis and Management of Asthma referenced a total of 1,654 studies to update the previous set of guidelines from 2004. The guidelines referenced 529 studies related to pharmacologic therapy for asthma, which included meta-analyses, systematic reviews of randomized controlled trials (RCTs), case control and cohort studies and non-analytic studies including case reports and case series. The guideline developers did not provide a breakdown of the specific number of randomized control trials (RCTs) and given the number of studies included in the systematic review we were not able to delineate all RCTs for each

recommendation. However, the evidence review table related to inhaled corticosteroid dosing, for example, included 34 RCTs or meta-analyses and systematic reviews of RCTs. This review is not comprehensive and represents only a portion of the research on this area.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall, the quality of the evidence regarding daily long-term asthma controller medication for patients who have persistent asthma assessment is high. The 34 RCTs referenced above included thousands of patients studied over long periods of time. The evidence supporting the recommendation of daily long-term asthma controller medication for patients with persistent asthma was graded Category A, which includes randomized controlled trials (RCTs), a rich body of data, evidence from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made and requires substantial numbers of studies involving substantial numbers of participants.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The studies included evidence-based guidelines with and without systematic reviews/ evaluations, economic evaluations of asthma medications, survey based research and retrospective studies. Research and studies consistently show that appropriate medication management could potentially prevent a significant proportion of asthma-related costs.

The evidence for daily long-term asthma control medication in patients with persistent asthma shows consistent benefit and high magnitude. The guidelines referenced 12 studies showing that the clinical effects of inhaled corticosteroids include reduction in severity of symptoms; improvement in asthma control and quality of life; improvement in peak expiratory flow and spirometry; diminished airway hyperresponsiveness; prevention of exacerbations; reduction in systemic corticosteroid courses, ED care, hospitalizations, and deaths due to asthma; and possibly the attenuation of loss of lung function in adults. Patients who have mild or moderate persistent asthma and are treated with ICS, compared to other single long-term control medications, demonstrate greater improvements in prebronchodilator forced expiratory volume (but not with long-term postbronchodilator forced expiratory volume); reduced airway hyperresponsiveness, symptom scores, exacerbation rates, and symptom frequency; as well as less use of supplemental SABA, fewer courses of oral systemic corticosteroids, and less use of hospitalization.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Benefits: Good control of asthma symptoms

• Improved quality of life

• Reduction in the frequency and severity of asthma exacerbations

• Fewer ED visits

Harms: Potential adverse effects of long-term medication use

The majority of research on harms relates to potential side-effects of asthma medications. However, the guidelines state that "ICSs are the most effective long-term therapy available for mild, moderate, or severe persistent asthma; in general, ICSs are well tolerated and safe at the recommended

dosages (Evidence A). The potential but small risk of adverse events from the use of ICS treatment is well balanced by their efficacy."

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

We searched for studies published in peer-reviewed journals specifically related to the correlation between asthma controller medication adherence and improved outcomes. One study found weak evidence of a link between 75% medication adherence and outcomes but three other studies found stronger associations between adherence and outcomes.

1a.8.2. Provide the citation and summary for each piece of evidence.

Yoon A, Crawford W, Sheikh J, Nakahiro R, Gong A and Schatz M. The HEDIS Medication Management for People with Asthma Measure is Not Related to Improved Asthma Outcomes. J Allergy Clin Immunol Pract 2015; 3:547-52. Summary: found that patients who achieved 75% compliance in 2012 did not have fewer hospitalizations or ED visits in 2013 compared to those who were not 75% compliant. They also found that patients who achieved 50% threshold in 2012 did not have fewer hospitalizations, but did have fewer ED visits in 2013, compared to those who were not 50% compliant. The study did not control for asthma history and risk of prior exacerbations, which could affect outcomes.

Stern L, Berman J, Lumry W, Katz L, Wang L, Rosenblatt L, et al. Medication compliance and disease exacerbation in patients with asthma: a retrospective study of managed care data. Ann Allergy Asthma Immunol 2006; 97:402-8. Summary: Based on a 75% medication possession ratio, more-compliant patients were less likely to experience exacerbation than less-compliant patients.

Lasmar L, Camargos P, Champs NS, Fonseca MT, Fontes MJ, Ibiapina C, et al. Adherence rate to inhaled corticosteroids and their impact on asthma control. Allergy 2009;64:784-9. Summary: In pediatric patients with asthma, there were statistically significant differences in adherence rates for maintaining or not maintaining the asthma control. Optimal asthma control entailed adherence rate higher than 80%. Strategies for reducing asthma morbidity should include a regular monitoring of adherence to inhaled steroids.

Williams LK, Pladevall M, Xi H, Peterson EL, Joseph C, Lafata JE, et al. Relationship between adherence to inhaled corticosteroids and poor outcomes among adults with asthma. J Allergy Clin Immunol 2004; 114:1288-93. Summary: Adherence to ICS was significantly and negatively correlated with the number of emergency department visits, the number of fills of an oral steroid, and the total days' supply of oral steroid. During the two-year study period, there were 80 asthma-related hospitalizations; an estimated 32 hospitalizations would have occurred were there no gaps in medication use.

Williams, L. K., E.L. Peterson, K. Wells, B.K. Ahmedani, R. Kumar, E.G. Buchard, V.K. Chowdhry, D. Favro, D.E. Lanfear, M. Pladevall. 2011. Quantifying the proportion of severe asthma exacerbations attributable to inhaled corticosteroid nonadherence. Journal of Allergy and Clinical Immunology, no 128.6 p. 1185-1191. Summary: Patients with asthma controller medication adherence rates of 75% or greater of the prescribed dose had a statistically significant reduction in asthma exacerbations compared to patients whose controller adherence was 25% or less.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
MMA_Evidence.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure assesses adherence to long-term asthma controller medications in patients with persistent asthma. The improvement in quality envisioned by the use of this measure is increasing adherence to long-term asthma controller medications in patients with persistent asthma. According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered “not well controlled” or “very poorly controlled” as defined by clinical practice guidelines (Stillman 2010). Increasing adherence to asthma controller medications can prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami 2009; (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007).

Akinbami, L.J., J.E. Moorman, P.L. Garbe, E.J. Sondik. 2009. Status of Childhood Asthma in the United States, 1980–2007. Pediatrics 123;S131-45. doi: 10.1542/peds.2008-2233C.

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf (November 19, 2015).

Stillman, L. 2010. Living with Asthma in New England: Results from the 2006 BRFSS and Call-back Survey. A report by the Asthma Regional Council of New England (February). http://www.hria.org/uploads/catalogerfiles/living-with-asthma-in-new-england/HRiA_Living_with_Asthma_BRFSS_2010.pdf (November 19, 2015).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included).*

This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.

The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year and product line (i.e. commercial, Medicaid, HMO and PPO).

The following data demonstrate the variation in the rate of medication adherence to asthma controller medications for children and adults with persistent asthma across health plans. In 2014 for the Medication Adherence 50% indicator, there was a 16 percentage point difference between commercial plans in the 10th percentile and commercial plans in the 90th percentile and 26 percentage point difference for Medicaid plans. Similarly in 2014 for the Medication Adherence 75% indicator, there was a 20 percentage point difference between plans in the 10th percentile and plans in the 90th percentile for commercial plans and 26 percentage point difference for Medicaid plans. These gaps in performance underscore the opportunity for improvement.

Medication Management for People with Asthma Numerator 1 (Medication Adherence 50%)

Commercial Rate Ages 5-64 (HMO and PPO Combined)

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 66% | 7% | 16% | 58% | 62% | 66% | 70% | 74% | 92% | 8%
2013 | 70% | 9% | 16% | 62% | 66% | 69% | 75% | 84% | 93% | 9%
2014 | 69% | 7% | 46% | 61% | 65% | 69% | 73% | 77% | 91% | 8%

Medicaid Rate Ages 5-64

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 51% | 9% | 23% | 41% | 45% | 51% | 56% | 62% | 90% | 11%
2013 | 54% | 10% | 25% | 44% | 48% | 54% | 59% | 67% | 94% | 11%
2014 | 54% |10% | 34% | 42% | 47% | 54% | 60% | 68% | 87% | 13%

Medication Management for People with Asthma Numerator 2 (Medication Adherence 75%)

Commercial Rate Ages 5-64 (HMO and PPO Combined)

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 43% | 8% | 8% | 34% | 39% | 42% | 47% | 51% | 82% | 8%
2013 | 48% | 11% | 7% | 37% | 40% | 45% | 52% | 66% | 82% | 12%
2014 | 45% | 8% | 12% | 35% | 40% | 45% | 50% | 55% | 72% | 10%

Medicaid Rate Ages 5-64

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 29% | 9% | 9% | 19% | 22% | 28% | 33% | 39% | 77% | 11%
2013 | 31% | 10% | 9% | 20% | 25% | 30% | 36% | 43% | 84% | 11%
2014 | 31% | 10% | 13% | 19% | 24% | 30% | 35% | 45% | 75% | 11%

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Commercial HMO

YEAR | N Plans | Mean Denominator Size per plan
2012 | 199 | 1,291
2013 | 200 | 1,169
2014 | 194 | 1,070

Commercial PPO

YEAR | N Plans | Mean Denominator Size per plan
2012 | 198 | 1,538
2013 | 196 | 1,566
2014 | 194 | 1,520

Medicaid HMO

YEAR | N Plans | Mean Denominator Size per plan
2012 | 154 | 1,640

2013 | 169 | 1,548
2014 | 192 | 1,628

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
HEDIS data are stratified by type of insurance (e.g. Commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.
Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts. Health Affairs 20(10): 1984-1991.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Although HEDIS measures are not stratified by race and ethnicity, researchers have explored disparities in asthma outcomes and in utilization to health care services among people with asthma. Children of low-income families experience more urgent care visits, hospitalizations and mortality due to asthma when compared to the general public (CDC 2009). Poor asthma outcomes in low-income children may be partly due to the barriers they face in accessing care for asthma, including access to medications. One study found that children with asthma from low-income families were less likely to have prescriptions filled and/or receive annual primary health examinations (Kim et al. 2009). The study also examined insurance coverage, showing that children without insurance coverage utilized primary health care services for asthma less often (Kim et al. 2009).

Centers for Disease Control and Prevention (CDC). Asthma: A Presentation of Asthma Management and Prevention, September 2009. http://www.cdc.gov/asthma/speakit/default.htm (November 19, 2015).

Kim, H., G.M. Kieckhefer, A.A. Greek, J.M. Joesch, N. Baydar. 2009. Health Care Utilization by Children With Asthma. Preventing Chronic Disease 6(1): A12.

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Asthma is one of the most prevalent chronic diseases. In 2010, 25.7 million Americans had asthma, including 7 million children, 15.6 million adults under 65 and 3.1 million adults 65 and older (Akinbami et al. 2012). Asthma has also become increasingly more common over the past decade, occurring in 7.3 percent of the population in 2001 compared to 8.4 percent in 2010 (Akinbami et al. 2012). Asthma is responsible for over 3,000 deaths in the U.S. annually (American Lung Association 2014) and accounted for over $50 billion spent on health care in the United States in 2007, an increase of almost $2 billion from 2002 (CDC 2011).

Appropriate medication adherence could ameliorate the severity of many asthma-related symptoms (Akinbami et al. 2009). According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered "not well controlled" or "very poorly controlled" as defined by clinical practice guidelines (Stillman 2010). Pharmacologic therapy is used to prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and reverse airflow obstruction (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007). Appropriate medication management could potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami et al. 2009). Indeed, several studies have found that higher medication adherence rates are associated with better outcomes; for instance, one study found that patients with asthma controller medication adherence rates of 75% or greater had fewer asthma exacerbations compared to patients with adherence rates of 25% or lower (Williams et al. 2011). The Asthma Regional Council has also stated that proper management could potentially save at least 25 percent of total asthma costs, or $5 billion, nationally by reducing health care costs (American Lung Association 2012).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Akinbami, L.J., J.E. Moorman, P.L. Garbe, E.J. Sondik. 2009. Status of Childhood Asthma in the United States, 1980–2007. Pediatrics 123;S131-45. doi: 10.1542/peds.2008-2233C.

Akinbami, L.J., J.E. Moorman, C. Bailey, H.S. Zahran, M. King, C.A. Johnson, X. Liu. 2012. "Trends in Asthma Prevalence, Health Care Use, and Mortality in the United States, 2001-2010." NCHS Data Brief, no. 94 (May). http://www.cdc.gov/nchs/data/databriefs/db94.pdf (November 19, 2015).

American Lung Association. 2012. Trends in Asthma Morbidity and Mortality. http://www.lung.org/finding-cures/our-research/trend-reports/asthma-trend-report.pdf (November 19, 2015).

American Lung Association. 2014. Asthma & Children Fact Sheet, September. http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/asthma/learn-about-asthma/asthma-children-facts-sheet.html (November 19, 2015).

Centers for Disease Control and Prevention (CDC). Vital Signs: Asthma in the US, May 2011. http://www.cdc.gov/VitalSigns/Asthma/index.html (November 19, 2015).

Stillman, L. 2010. Living with Asthma in New England: Results from the 2006 BRFSS and Call-back Survey. A report by the Asthma Regional Council of New England (February). http://hria.org/resources/reports/asthma/living-with-asthma-in-new-england.html (November 19, 2015).

Williams, L. K., E.L. Peterson, K. Wells, B.K. Ahmedani, R. Kumar, E.G. Buchard, V.K. Chowdhry, D. Favro, D.E. Lanfear, M. Pladevall. 2011. Quantifying the proportion of severe asthma exacerbations attributable to inhaled corticosteroid nonadherence. Journal of Allergy and Clinical Immunology, no 128.6 p. 1185-1191

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be***

evaluated against the remaining criteria.

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 1799_MMA_Value_Sets.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no significant changes to the measure specification since the last endorsement maintenance completed on January 31, 2012.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Numerator 1 (Medication Adherence 50%): The number of patients who achieved a PDC* of at least 50% for their asthma controller medications during the measurement year. A higher rate is better.

Numerator 2 (Medication Adherence 75%): The number of patients who achieved a PDC* of at least 75% for their asthma controller medications during the measurement year. A higher rate is better.

*PDC is the proportion of days covered by at least one asthma controller medication prescription, divided by the number of days in the treatment period. The treatment period is the period of time beginning on the earliest prescription dispensing date for any asthma controller medication during the measurement year through the last day of the measurement year.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: 12 month period (the measurement year)
Denominator: 24 month period (the measurement year and the year prior)
Exclusions: lookback through the patient's history through the last day of the measurement year

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Follow the steps below to identify numerator compliance.

Step 1: Identify the Index Prescription Start Date*. The Index Prescription Start Date is the earliest dispensing event for any asthma

controller medication (refer to MMA-B Asthma Controller Medications) during the measurement year.

Step 2: To determine the treatment period, calculate the number of days beginning on the Index Prescription Start Date through the end of the measurement year.

Step 3: Count the days covered by at least one prescription for an asthma controller medication (refer to MMA-B Asthma Controller Medications) during the treatment period. To ensure that days supply that extends beyond the measurement year is not counted, subtract any days supply that extends beyond the end of the of the measurement year (e.g., December 31).

Step 4: Calculate the patient's Proportion of Days Covered using the following equation. Round (using the .5 rule) to two decimal places.
(Total Days Covered by a Controller Medication in the Treatment Period (Step 3)
/Total Days in Treatment Period (Step 2))

Numerator 1 (Medication Adherence 50%): Sum the number of patients whose Proportion of Days Covered is > or =50% for their treatment period.

Numerator 2 (Medication Adherence 75%): Sum the number of patients whose Proportion of Days Covered is > or =75% for their treatment period

MMA-B: Asthma Controller Medications:
Antiasthmatic combinations: dyphylline-guaifenesin, guaifenesin-theophylline

Antibody inhibitor: omalizumab

Inhaled steroid combinations: budesonide-formoterol, fluticasone-salmeterol, mometasone-formoterol

Inhaled corticosteroids: beclomethasone, budesonide, ciclesonide, flunisolide, fluticasone CFC free, mometasone,

Leukotriene modifiers: montelukast, zafirlukast, zileuton

Mast cell stabilizers: cromolyn

Methylxanthines: aminophylline, dyphylline, theophylline

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
All patients 5–64 years of age as of December 31 of the measurement year who have persistent asthma by meeting at least one of the following criteria during both the measurement year and the year prior to the measurement year:
• At least one emergency department visit with asthma as the principal diagnosis
• At least one acute inpatient claim/encounter with asthma as the principal diagnosis
• At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits.
• At least four asthma medication dispensing events

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The eligible population for the denominator is defined by following the series of steps below:

Step 1: Identify patients as having persistent asthma who met at least one of the following criteria during both the measurement year and the year prior to the measurement year. Criteria need not be the same across both years.
• At least one ED visit (refer to codes in ED Value Set) with asthma as the principal diagnosis (refer to codes in Asthma Value Set).
• At least one acute inpatient claim/encounter (refer to codes in Acute Inpatient Value Set) with asthma as the principal diagnosis

(refer to codes in Asthma Value Set).
• At least four outpatient visits (refer to codes in Outpatient Value Set) or observation visits (refer to codes in Observation Value Set) on different dates of service, with any diagnosis of asthma (refer to codes in Asthma Value Set) AND at least two asthma medication dispensing events (see MMA-A). Visit type need not be the same for the four visits.
• At least four asthma medication dispensing events (see MMA-A)

Step 2: A patient identified as having persistent asthma because of at least four asthma medication dispensing events, where leukotriene modifiers or antibody inhibitors were the sole asthma medication dispensed in that year, must also have at least one diagnosis of asthma (refer to codes in Asthma Value Set), in any setting, in the same year as the leukotriene modifier or antibody inhibitor (i.e., measurement year or year prior to the measurement year).

See attached value set Excel document for the following value sets:
- ED Value Set
- Asthma Value Set
- Acute Inpatient Value Set
- Outpatient Value Set
- Observation Value Set

MMA-A: Asthma Medications
Antiasthmatic combinations: dyphylline-guaifenesin; guaifenesin-theophylline

Antibody inhibitor: omalizumab

Inhaled steroid combinations: budesonide-formoterol; fluticasone-salmeterol; Mometasone-formoterol

Inhaled corticosteroids: beclomethasone; budesonide; ciclesonide; flunisolide; fluticasone CFC free; mometasone

Leukotriene modifiers: montelukast; zafirlukast; zileuton

Mast cell stabilizers: cromolyn

Methylxanthines: aminophylline; dyphylline; theophylline

Short-acting, inhaled beta-2 Agonists: albuterol; levalbuterol; metaproterenol; pirbuterol

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
1) Exclude patients who had any of the following diagnoses any time during the patient's history through the end of the measurement year (e.g., December 31):
-COPD
-Emphysema
-Obstructive Chronic Bronchitis
-Chronic Respiratory Conditions Due To Fumes/Vapors
-Cystic Fibrosis
-Acute Respiratory Failure

2) Exclude any patients who had no asthma controller medications dispensed during the measurement year.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1) Exclude patients who had any diagnosis of Emphysema (refer to codes in Emphysema Value Set or Other Emphysema Value Set), COPD (refer to codes in COPD Value Set), Chronic Bronchitis (refer to codes in Obstructive Chronic Bronchitis Value Set), Chronic Respiratory Conditions Due To Fumes/Vapors (refer to codes in Chronic Respiratory Conditions Due to Fumes/Vapors Value Set), Cystic Fibrosis (refer to codes in Cystic Fibrosis Value Set) or Acute Respiratory Failure (refer to codes in Acute Respiratory Failure Value Set) any time during the patient's history through the end of the measurement year (e.g., December 31).

2) Exclude any patients who had no asthma controller medications (see MMA-B) dispensed during the measurement year.

See attached value set Excel document for the following value sets:
- Emphysema Value Set
– Other Emphysema Value Set
– COPD Value Set
– Obstructive Chronic Bronchitis Value Set
– Chronic Respiratory Conditions Due to Fumes/Vapors Value Set
– Cystic Fibrosis Value Set
– Acute Respiratory Failure Value Set

MMA-B: Asthma Controller Medications:
Antiasthmatic combinations: dyphylline-guaifenesin, guaifenesin-theophylline

Antibody inhibitor: omalizumab

Inhaled steroid combinations: budesonide-formoterol, fluticasone-salmeterol, mometasone-formoterol

Inhaled corticosteroids: beclomethasone, budesonide, ciclesonide, flunisolide, fluticasone CFC free, mometasone

Leukotriene modifiers: montelukast, zafirlukast, zileuton

Mast cell stabilizers: cromolyn

Methylxanthines: aminophylline, dyphylline, theophylline

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Four age stratifications and a total rate are reported for this measure. Age for each stratum is based on the patient's age as of the end of the Measurement Year (e.g., December 31).
1) 5–11 years
2) 12–18 years
3) 19-50 years
4) 51-64 years
5) Total (5-64 years)

The age strata align with both clinical practice guidelines and reporting requirements for child health quality improvement programs. Clinical guidelines specify appropriate age cohorts for measuring use of asthma medications as 5–11 years of age and 12–50 years of age, to account for the differences in medication regimens for children compared to adolescents and adults. Implementation requires further stratification of the age ranges to enable creation of comparable cohorts that align with child health populations.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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.)*
Refer to items S.6 (Numerator details), S.9 (Denominator details), S.11 (Denominator exclusions details) and S.2b (Data Dictionary) for tables.

This measure determines the number of days covered with a controller medication based on information available from the published NDC codes to calculate adherence to asthma medications. The measure calculation is detailed in the steps listed below:

Step 1: Determine the eligible population: Identify patients 5–64 years of age as of December 31 of the measurement year as having persistent asthma who met at least one of the following criteria during both the measurement year and the year prior to the measurement year. Criteria need not be the same across both year:
a) At least one ED visit with asthma as the principal diagnosis; or
b) At least one acute inpatient claim/encounter with asthma as the principal diagnosis; or
c) At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits; or
d) At least four asthma medication dispensing events*

*A patient identified as having persistent asthma because of at least four asthma medication dispensing events where leukotriene modifiers or antibody inhibitors were the sole asthma medication dispensed in that year, must also have at least one diagnosis of asthma, in any setting, in the same year as the leukotriene modifier or antibody inhibitor (i.e., measurement year or year prior to the measurement year).

Step 2: Determine denominator exclusions:
a) Exclude patients who had any diagnosis of Emphysema, COPD, Chronic Bronchitis, Chronic Respiratory Conditions Due to Fumes/Vapors, Cystic Fibrosis or Acute Respiratory Failure any time during the patient's history through the end of the measurement year
b) Exclude patients who had no asthma controller medications dispensed during the measurement year.

Step 3: Determine numerator:
a) Identify the Index Prescription Start Date. The Index Prescription Start Date is the earliest dispensing event for any asthma controller medication during the measurement year.
b) To determine the treatment period, calculate the number of days beginning on the Index Prescription Start Date through the end of the measurement year.
c) Count the days covered by at least one prescription for an asthma controller medication during the treatment period. To ensure that days supply that extends beyond the measurement year is not counted, subtract any days supply that extends beyond the end of the of the measurement year (e.g., December 31).
d) Calculate the patient's Proportion of Days Covered using the following equation. Round (using the .5 rule) to two decimal places:
(Total Days Covered by a Controller Medication in the Treatment Period/Total Days in Treatment Period)
e) Calculate Numerator 1: Sum the number of patients whose Proportion of Days Covered is > or =50% for their treatment period.
f) Calculate Numerator 2: Sum the number of patients whose Proportion of Days Covered is > or =75% for their treatment period

Step 4: Calculate two rates:
a) Number of patients whose PDC is > or =50% for their treatment period/Denominator
b) Number of patients whose PDC is > or =75% for their treatment period/Denominator

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
MMA_Testing.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**: 1799**
Measure Title: **Medication Management for People with Asthma**
Date of Submission: **12/14/2015**
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
N/A

1.3. What are the dates of the data used in testing? Click here to enter date range
Initial testing: During measure development, we conducted a comprehensive field test to assess feasibility of data collection and validity of the numerator, denominator and exclusions. This field test used data from measurement year 2009, which included health plan data spanning January 1, 2008 through December 31, 2009.

Systematic evaluation of face validity: The measure was tested for face validity throughout measure development from 2010 to 2012. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2012.

2015 Update: We assessed measure score reliability and construct validity using data from all health plans that submitted HEDIS data to NCQA for this measure in 2015, which used data for measurement year 2014. Measurement year 2014 required two years' worth of health plan data from January 1, 2013 through December 31, 2014.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the*

analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample)

Initial testing: To assess feasibility of data collection and validity of measure denominator, exclusions and numerator, 9 health plans (7 Commercial and 7 Medicaid plans) provided individual member-level data to NCQA for analysis. These plans were selected because they had the resources to generate the files, had sufficient sample of members with persistent asthma for analysis, and willingness to provide the data. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity: Throughout the entire measure development process from 2010-2012, the measure was tested for face validity using panels of experts with specific clinical, methodologic and operational expertise. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2012. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliation of expert panel:

1) NCQA's Respiratory Measurement Advisory Panel (RMAP) is comprised of 10 experts (8 physicians, 1 pharmacist and 1 researcher) in clinical pulmonary care, including health care providers and policy makers.

2) NCQA's Technical Measurement Advisory Panel is a 12-member panel representing health plans methodologists, clinicians and HEDIS auditors.

3) NCQA's Committee on Performance Measurement (CPM) oversees the HEDIS measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 17 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

4) [2015 Update] NCQA's HEDIS Expert Coding Panel reviewed and provided feedback on the vocabularies and definitions found in the values sets used to identify each measure component as well as the more recent mapping of ICD-9 codes to ICD-10 codes.

In 2011, the draft measure was posted for public comment, a 30-day period of review that allowed interested parties to offer feedback to NCQA about the measure. Stakeholders from various types of organizations submitted 43 comments on the measure.

2015 Update: Measure score reliability and construct validity was calculated from the 388 Commercial health plans and 192 Medicaid health plans that submitted data on this measure to HEDIS in 2015. The plans were geographically diverse and varied in size.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Patient sample for initial measure field testing: We collected data from 9 plans (7 Commercial and 7 Medicaid plans). Below is a description of the sample. It includes the number of health plans that provided data for the measurement year 2009 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial	7	3,920
Medicaid	7	2,577

2015 Update: Measure score reliability and construct validity testing: In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the number of health plans that submitted data for this measure to HEDIS for measurement year 2014 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan

Commercial	388	514
Medicaid	192	921

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

During measure development, we conducted a comprehensive field test to assess feasibility of data collection, reliability and validity of the numerator, denominator and exclusions using data submitted by 9 plans (7 Commercial and 7 Medicaid plans). The field test used the measurement year 2009 that included data from January 1, 2008 through December 31, 2009.

Face validity was demonstrated through a systematic assessment of face validity during measure development and at regular intervals since then. Per NQF instructions we have described the composition of the technical expert panel which assessed face validity in the data sample questions above.

2015 Update: The measure underwent additional analyses to assess measure score reliability (tested using a beta-binomial calculation) and construct validity (tested using Pearson's correlations of similar measures). These analyses included all of the health plans (388 Commercial and 192 Medicaid) that submitted data for this measure to HEDIS for measurement year 2014 (using data from January 1, 2013 through December 31, 2014).

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Measure performance results are stratified by commercial and Medicaid health plans and by age (5-11 years; 12-18 years; 19-50 years; 51-64 years).

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual

accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Beta-Binomial Statistic for Each Measure Rate:

	Commercial			Medicaid		
	Median	Overall	10th-90th	Median	Overall	10th-90th
Numerator 1 (Medication Adherence 50%)	0.90	0.84	0.58-0.98	0.97	0.93	0.83-0.99
Numerator 2 (Medication Adherence 75%)	0.92	0.87	0.65-0.99	0.98	0.94	0.85-0.99

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Interpretation of measure score reliability testing:
Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. Asthma beta binomial testing suggests that both indicators within this measure have good reliability. The 10-90th percentile distribution of health plan level-reliability on the rates in this measure show the vast majority of health plans exceeded the minimally accepted threshold of 0.7, and the majority of plans exceeded 0.8. Strong reliability is demonstrated since majority of variances is due to signal and not to noise.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

METHOD OF ASSESSING FACE VALIDITY

2012 Submission Form [Testing Data]: NCQA tested the measure results for face validity using a panel of stakeholders with relevant clinical expertise and research and measurement, experience. This panel included representatives from key stakeholder groups, including the CDC, pulmonologists, provider and deliver organizations and researchers (See list of members for the Respiratory Advisory Panel (RMAP) under section Ad.1). RMAP experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2015 Update: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature

review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

The asthma medication management measure was developed to address adherence to long-term asthma controller medications in patients with persistent asthma in 2010. NCQA and the Respiratory Measurement Advisory Panel worked together to develop the most appropriate measure for assessing asthma medication adherence.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

The asthma medication management measure was written and field-tested in 2010. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in 2011.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

The asthma medication management measure was released for Public Comment in 2011 prior to publication in HEDIS. We received and responded to 43 comments on this measure. The CPM recommended moving this measure to first year data collection by a majority vote.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

The asthma medication management measure was introduced to HEDIS in 2011. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting with a majority vote.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publically reported and may be used for scoring in accreditation.

The asthma medication management measure has been publicly reported in HEDIS since 2012.

STEP 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

The clinical guideline recommendations for medication management of asthma have not changed since the measure was developed in 2010; therefore, we have not made any significant changes to the medication management measure since it was last endorsed on January 31, 2012.

Expert Participation

This measure was tested for face validity with input from three expert panels. Guidelines from the National Heart, Lung and Blood Institute/National Asthma Education and Prevention Program in 2007 were also a strong authoritative source in applying the evidence for medication management for people with asthma measure.

We list an overview of each panel here. Please refer to Ad.1 in the submission form for the names and affiliation of experts in each panel.

1) Respiratory Measurement Advisory Panel includes ten members (eight physicians, one pharmacist and a researcher) with expertise in respiratory care and quality measurement.
2) The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.
3) NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

ICD-10 CONVERSION:
In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process
1. NCQA first identified value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM)to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.
3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented to for expert review and feedback.
4. NCQA RMAP clinical review: Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.
5. New value sets containing ICD-10 code recommendations were for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under **Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development**.

METHOD OF TESTING CONSTRUCT VALIDITY

2015 Update: We tested for construct validity by exploring whether the indicators within this measure were correlated with other similar measures of respiratory care. We hypothesized that organizations that perform well on the two indicators should perform well on other similar HEDIS measures. To test these correlations we used a Person correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A

value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

For this measure, we specifically hypothesized:
1) The medication adherence of 50% indicator will be positively correlated with the medication adherence of 75% indicator
2) The medication adherence of 50% indicator will be positively correlated with the asthma medication ratio measure (a measure of good asthma control)
3) The medication adherence of 75% indicator will be positively correlated with the asthma medication ratio measure

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

RESULTS OF FACE VALIDITY ASSESSMENT

2015 Submission Form [Testing Data]: For the initial field test in 2010, we calculated performance rates on the measure indicators stratified by age group and discussed the validity of the performance results with the expert panels (Respiratory Measurement Advisory Panel and the Committee on Performance Measurement). The expert panels agreed that the performance on the 50% and 75% thresholds for medication adherence were accurate representations of quality performance and distinguished performance among health plans.

2010 Field Test: Performance Rates on the Medication Management for People with Asthma Measure*

	Age Group	Den	Med. Adherence ≥ 50%		Med. Adherence ≥ 75%	
			Numerator	Rate	Numerator	Rate
Commercial	Ages 5-11	5,685	3,000	52.8%	1,707	30.0%
	Ages 12-50	19,289	10,237	53.1%	5,965	30.9%
	Ages 51-64	10,963	6,874	62.7%	4,629	42.2%
	Total (Ages 5-64)	35,937	20,111	56.0%	12,301	34.2%
Medicaid	Ages 5-11	8,397	3,241	38.6%	1,777	21.2%
	Ages 12-50	11,982	4,230	35.3%	2,398	20.0%
	Ages 51-64	1,570	731	46.6%	531	33.8%
	Total (Ages 5-64)	21,949	8,202	37.4%	4,706	21.4%

**Includes data submitted by 7 Commercial plans and 7 Medicaid plans using measurement year 2009*

We assessed validity of the denominator criteria to ensure the measure is capturing people with persistent asthma. Entry into the eligible population for persistent asthma requires a combination of multiple outpatient encounters and diagnoses. Approximately 90% of commercial members and 88% of Medicaid members were included in the eligible population by having at least four asthma medication dispensing events in the measurement year and the year prior. The remaining 10-12% of members had at least one ED visit with a diagnosis of asthma, one inpatient visit with a diagnosis of asthma, or four outpatient visits with a diagnoses of asthma plus two asthma medication dispensing events in the measurement year and the year prior.

We examined whether encounters could be linked to the same event and therefore do not accurately capture a population with persistent asthma. Using the field test dataset, NCQA examined the different scenarios where encounters were less than 14 days apart (a standard HEDIS time frame for linked encounters) to determine the effect on the measure's eligible population. Section 2b3.3 details the results of this additional analysis revealing the proportion of the population that would potentially excluded from the EP as a result of the additional criterion of <14 days between encounters. The next table details the proportion of the population that would potentially be excluded from the EP as a result of the additional criterion of <14 days between encounters. For this table, "Eligible Population" (EP) refers only to those members who satisfied the "Combination" eligibility criterion of at least four outpatient encounters and at least two prescription events.

Proportion of the Eligible Population affected by a ≥14 Day rule.

	Commercial				Medicaid			
		Do Not Qualify				Do Not Qualify		
Age Group	EP	N	% of EP	% of total EP	EP	N	% of EP	% of total EP
5-11	293	77	26.3%	1.3%	78	17	21.8%	0.2%
12-50	634	137	21.6%	0.6%	88	16	18.2%	0.1%
51-64	715	104	14.5%	0.6%	14	0	0.0%	0.0%
Total 1 (5-50)	927	214	23.1%	0.7%	166	33	19.9%	0.1%
Total 2 (5-64)	1,854	428	23.1%	0.9%	180	33	18.3%	0.1%

Another concern when measuring management for plan-to-plan comparison is ensuring that the majority of index prescriptions occur at a point within the measurement year (Q1, Q2, Q3, & Q4) that objectively monitors adherence without any type of adjustment. It addresses the question: Is the prescription utilization stable for this population and, if so, is the administrative data capturing index prescription start dates (IPSDs) sufficiently early in the measurement year to adequately measure medication management. The following table outlines the percentage of index prescriptions occurring in each quarter of the measurement year by cohort. The table presents the percentage of index prescriptions dispensed to members of the entire Eligible Population after comorbidity exclusions have been applied.

Timing of Index Prescription (by Age Group and line of business)

Product	Age	Q1	Q2	Q3	Q4
Commercial	5-11	67.2%	16.1%	6.9%	5.3%
	12-50	65.1%	14.3%	6.1%	4.3%
	51-64	72.7%	12.7%	4.0%	2.5%
Medicaid	5-11	62.7%	15.0%	6.1%	4.9%
	12-50	55.5%	12.9%	5.8%	2.1%
	51-64	58.9%	6.3%	2.8%	2.1%

The expert panels agreed that the denominator as specified was valid in identifying people with persistent asthma.

RESULTS OF CONSTRUCT VALIDITY TESTING

2015 Update: The results indicated that the asthma measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized. The level of correlations among 2 out of the 3 indicators for both Commercial and Medicaid plans ranged from moderate to strong (the correlation coefficients were higher than 0.3).

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Commercial Plans)*

Asthma Measure	Pearson Correlation Coefficients		
	Medication Adherence 50%	Medication Adherence 75%	Asthma Medication Ratio
Medication Adherence 50%	-	0.9	0.3
Medication Adherence 75%	-	-	0.2

**Includes data submitted by 388 Commercial plans to HEDIS for these measures for measurement year 2014*

Note: All correlations are significant at p<.05

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Medicaid Plans)*

Asthma Measure	Pearson Correlation Coefficients		
	Medication Adherence 50%	Medication Adherence 75%	Asthma Medication Ratio

Medication Adherence 50%	-	1.0	0.3
Medication Adherence 75%	-	-	0.2

**Includes data submitted by 192 Medicaid plans to HEDIS for these measures for measurement year 2014*

Note: All correlations are significant at p<.05

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

SYSTEMATIC ASSESSMENT OF FACE VALIDITY

2015 Update: The current measure for children and adults ages 5-64 was deemed to have the desirable attributes of a HEDIS measure in 2011 (relevance, scientific soundness, and feasibility). The technical expert panels showed good agreement that the measure as specified accurately identifies patients with persistent asthma and differentiates quality across providers. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since then. Our interpretation of these results is that this measure has sufficient face validity.

CONSTRUCT VALIDITY

2015 Update: Coefficients with absolute value of less than 0.2 are generally considered indicative of weak associations whereas absolute values of 0.2 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that the asthma measures are correlated with each other, suggesting they represent the same underlying quality construct of asthma quality of care.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2012 Submission Form [Testing Data]: The measure is intended to assess patients with persistent asthma whose asthma is being controlled through long-term asthma controller medications and to align with the clinical guideline recommendations for medication management of persistent asthma. The measure is not intended to capture people with intermittent or seasonal asthma or those who have non-asthma respiratory conditions. To ensure the measure captures people with persistent asthma only, members must meet one of four criterion (including asthma medication dispensing events and/or encounters with an asthma diagnosis) in both the measurement year and the year prior. The measure also excludes people with a diagnosis for a specific clinical condition (COPD, emphysema, obstructive chronic bronchitis, cystic fibrosis and acute respiratory failure). During measure development in 2010, exclusions were tested using data from 7 commercial and 7 Medicaid health plans to determine the impact each clinical condition had on the measure's performance. We calculated the percent of people excluded from the denominator (i.e., the percent who had at least 1 exclusion condition) and the total percent of people excluded from the denominator for each age cohort.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2012 Submission Form [Testing Data]: A total of 25% of Commercial members and 18% of Medicaid members were excluded from the measure. A higher percentage of people ages 51-64 had at least 1 measure exclusion compared to children or adults ages 5-50.

Impact of Co-morbidity Exclusions on the Eligible Population*

	Age	EP	Percent of Eligible Population Excluded for a Co-Morbidity

	Range		At least 1 exclusion	COPD	Chronic Bronchitis	Emphysema	Cystic Fibrosis	Acute Respiratory Syndrome
Comm. N=7 plans	5 - 11	6,031	5.7%	3.7%	1.3%	0.3%	0.8%	1.0%
	12 - 50	22,855	16.2%	14.2%	4.1%	0.9%	0.6%	1.3%
	51 - 64	18,154	41.5%	39.6%	15.6%	6.4%	0.1%	3.5%
	Total (5-64 years)	47,040	24.6%	22.6%	8.2%	3.0%	0.4%	2.1%
Medicaid N=7 plans	5 - 11	8,614	3.8%	2.7%	0.4%	0.1%	0.4%	0.6%
	12 - 50	14,337	18.3%	16.6%	3.4%	1.3%	0.4%	2.6%
	51 - 64	4,432	45.2%	43.7%	13.8%	6.7%	0.2%	7.4%
	Total (5-64 years)	27,383	18.1%	16.6%	4.2%	1.8%	0.3%	2.7%

EP= total eligible population (number of patients) across all health plans prior to exclusions
**Includes data submitted by 7 Commercial plans and 7 Medicaid plans using measurement year 2009*

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2015 Update: The denominator criteria and the exclusions in this measure are intended to focus the measure on children and adults who have persistent asthma rather than concomitant diagnoses of asthma and COPD or chronic bronchitis. The higher percentage of adults ages 51-64 being excluded from the measure was expected, as they have a higher prevalence of conditions such as COPD or chronic bronchitis. The measure exclusions are needed in order to: 1) optimize the specificity of the denominator to only include those with persistent asthma and to keep the measure aligned with the clinical guideline recommendations; and to 2) display performance results that truly reflect appropriate care for this cohort of patients.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by ___risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2012 Submission Form [Testing Data]: Nine health plans covering a variety of geographic areas within the United States were asked to provide complete administrative data file consisting of any member in their commercial and Medicaid product lines for anyone that had a diagnosis code for asthma during the calendar years of 2009-2010. The complete member-level administrative file used for analysis included a total of more than 82,000 health plan members with persistent asthma. Specific calculations involve average performance rate, distribution (percentiles), 95% confidence interval of average rate across the respective health plans per by product line.

2015 Update: To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measured entities. However the method can be used for comparison of any two measured entities.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2012 Submission Form [Testing Data]: Distribution of plan performance for the field test data set by each product line (commercial and Medicaid).

	Product Line	Ave Rate	Lower 95% CI	Upper 95% CI	Standard Deviation	Min Rate	Max Rate	10th	25th	50th	75th	90th
50% PDC	Comm.	0.53095	0.45935	0.60255	0.07742	0.37500	0.60479	0.37500	0.50142	0.55019	0.58541	0.60479
	Medicaid	0.39322	0.33552	0.45093	0.06239	0.31546	0.50000	0.31546	0.34541	0.38106	0.43912	0.50000
75% PDC	Comm.	0.32816	0.26624	0.39008	0.06696	0.20052	0.40562	0.20052	0.30175	0.33483	0.38860	0.40562
	Medicaid	0.23024	0.18622	0.27426	0.04760	0.15097	0.29710	0.15097	0.20211	0.22705	0.27473	0.29710

2015 Update: HEDIS 2014 Variation in Performance across Health Plans, Medication Adherence ≥50%*

	Ages	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Comm.	Ages 5-11	263	63	9	51	57	63	68	73	11	<0.001
	Ages 12-18	225	59	8	50	55	59	64	69	9	<0.001
	Ages 19-50	587	67	7	58	63	67	72	77	9	<0.001
	Ages 51-64	477	77	7	69	73	77	81	85	8	0.001
	Total (Ages 5-64)	1,295	69	9	61	65	69	73	77	7	<0.001
Medicaid	Ages 5-11	800	51	11	38	44	50	58	64	14	<0.001
	Ages 12-18	523	48	9	38	42	47	52	60	10	<0.001
	Ages 19-50	376	59	8	48	56	59	63	68	7	<0.001
	Ages 51-64	178	72	7	62	67	73	76	79	9	<0.001
	Total (Ages 5-64)	1,627	54	10	42	47	54	60	68	12	<0.001

**Includes data submitted by 388 Commercial plans and 192 Medicaid plans to HEDIS for this measure for measurement year 2014*
EP: Eligible Population, the average denominator size across all plans submitting 2014 HEDIS data
IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

HEDIS 2014 Variation in Performance across Health Plans, Medication Adherence ≥75%*

	Ages	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Comm.	Ages 5-11	263	38	9	27	32	38	43	49	11	<0.001
	Ages 12-18	225	35	8	26	29	35	39	44	10	0.015
	Ages 19-50	587	43	9	33	37	43	47	53	10	<0.001
	Ages 51-64	477	55	9	44	50	56	61	65	11	<0.001
	Total (Ages 5-64)	1,295	45	8	35	40	46	50	55	10	<0.001
Medicaid	Ages 5-11	800	27	10	16	20	25	33	40	13	<0.001
	Ages 12-18	523	24	8	15	18	23	29	36	11	<0.001
	Ages 19-50	376	36	9	23	31	35	41	47	10	0.004
	Ages 51-64	178	48	10	37	42	49	54	59	12	<0.001
	Total (Ages 5-64)	1,627	31	10	19	24	30	35	45	11	<0.001

*Includes data submitted by 388 Commercial plans and 192 Medicaid plans to HEDIS for this measure for measurement year 2014
EP: Eligible Population, the average denominator size across all plans submitting 2014 HEDIS data
IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.e., *what do the results mean in terms of statistical and meaningful differences?*)
2015 Update: The results above indicate there is a 7-14% gap in performance between the 25th and 75th percentile performing plans across the different age ranges and product lines. For all product lines and age groups the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance is for Medicaid plans on the Medication Adherence 75% indicator for children ages 5-11, which shows a 14 percentage point gap between 25th and 75th percentile plans.

To put these meaningful differences in performance into context, we estimated that on average 179 additional members per plan would have medication adherence ≥75% if plans in the 25th percentile performed as well as plans in the 75th percentile. This estimate is based on the average health plan eligible population.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.e., *what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
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 all 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 followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will 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

In addition to the HEDIS Audit, NCQA provides a system to allow “real-time” feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Rating

	http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Medicaid Children's Core Set https://www.medicaid.gov/medicaid-chip-program-information/by-topics/quality-of-care/downloads/medicaid-and-chip-child-core-set-manual.pdf Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Regulatory and Accreditation Programs NCQA Health Plan Accreditation http://www.ncqa.org/tabid/123/Default.aspx Health Insurance Marketplace: Quality Rating System (2016) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/2015-QRS-Measure-Technical-Specifications.pdf Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.

HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. [REPLACE or ADD as appropriate, 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives.] Health plans are scored based on performance compared to benchmarks.

QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

MEDICAID CHILD CORE SET: These are a core set of health quality measures for Medicaid and CHIP-enrolled children. The Medicaid Child Core Set was identified by the Centers for Medicare & Medicaid (CMS) in partnership with the Agency for HealthCare Research and Quality (AHRQ). The data collected from these measures will help CMS to better understand the quality of health care that children enrolled in Medicaid and CHIP programs receive nationally. The initial core set was published in February 2011. CHIPRA required the Secretary to publish annual changes to the Child Core Set beginning in January 2013, and an annual Secretary's report on the quality of care for children enrolled in Medicaid and CHIP is released every September summarizing state-specific and national information on the quality of health care furnished to children enrolled in Medicaid and CHIP.

CMS HEALTH INSURANCE MARKET QUALITY RATING SYSTEM: This measure is used in the CMS developed, Quality Reporting Rating System (QRS) set of measures. The QRS measure set consists of measures that address areas of clinical quality management; enrollee experience; and plan efficiency, affordability and management. The measure set includes a subset of NCQA's HEDIS measures and one PQA measure.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

From 2012-2014, the Medication Adherence 50% indicator showed slight improvement (approximately 3 percentage points) across Commercial and Medicaid health plans (see section 1b.2 for summary of data from health plans). There was also improvement in performance for Commercial and Medicaid plans at the 90th percentile (+4 percentage points for Commercial plans and +6 percentage points for Medicaid plans). These data are nationally representative.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
More Medicaid plans reported the measure in 2014 compared to 2013 and 2012, which may help explain why the performance rates did not substantially improve. There is hope that with increasing attention to this measure in public reporting programs such as the Medicaid Child Core Set and in accreditation programs such as the Health Insurance Marketplace: Quality Rating System, performance rates on the Medication Adherence 75% indicator will begin to improve across more plans and performance rates on the Medication Adherence 50% indicator will continue to improve across plans.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no identified unintended consequences for this measure during testing or since implementation.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0047 : Asthma: Pharmacologic Therapy for Persistent Asthma
0548 : Suboptimal Asthma Control (SAC) and Absence of Controller Therapy (ACT)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
0047 is a physician-level measure that assesses whether a patient was prescribed medication at least once during the measurement year, while our measure assesses patient adherence to asthma controller medications throughout the measurement year. 0548 is a health plan-level measure that assesses two rates of poor asthma control that indicate over-utilization of rescue medication and need for additional therapeutic intervention; meanwhile our measure assesses patient adherence to asthma controller medications during the measurement year. There is no impact on interpretability or added burden of data collection because the focus of each measure is different and the data for each measure is collected from different data sources by different entities.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955--

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Respiratory Measurement Advisory Panel (RMAP) Members:
David Au, MD, MS, (CHAIR) Associate Prof. of Medicine/Investigator HSRD, Department of Veterans Affairs
Kurt Elward, MD, MPH, Clinical Professor of Family Medicine, Virginia Commonwealth University
Laura Feemster, MD, MS, Assistant Professor of Medicine, University of Washington Medical Center, VA HSR&D
Anne Fuhlbrigge, MD, Clinical Director, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School
Min Joo, MD, MPH, FCCP, Assistant Professor of Medicine, University of Illinois at Chicago/ Jesse Brown VA Medical Center
Christine Joseph, PhD, MPH, BSc, Associate Director of Research, Epidemiologist, Henry Ford Health System
Jerry Krishnan, MD, PhD, Prof. of Medicine & Public Health, Director of Population Health Sciences, AVP, Office of the VP for Health Affairs, University of Illinois Hospital & Health Sciences System
Todd Lee, PharmD, PhD, Primary: Senior Investigator, Secondary: Associate Professor, University of Illinois at Chicago
Allan Luskin, MD, Physician Pulmonologist, Healthy Airways
Richard O´Connor, MD, Director, Dept. of Quality Management, Allergist/Immunologist, Sharp Rees-Stealy Medical Group

Committee on Performance Measurement (CPM) Members:
Peter Bach, MD, Director, Health Policy Center, Memorial Sloan Kettering Cancer Center
Bruce Bagley, MD, FAAFP, Senior Advisor to the Professional Satisfaction and Practice Sustainability effort at the American Medical Association
Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna
Patrick Conway, MD, MSC, Chief Medical Officer and Deputy Administrator, Centers for Medicare and Medicaid Services
Jonathan D. Darer, MD, MPH, Chief Innovation Officer, Geisinger Health System
Helen Darling, Strategic Advisor, National Business Group of Health
Rebekah Gee, MD, MPH, FACOG, Assistant Professor, LSUHSC
Foster Gesten, MD, FACP, New York State Department of Health
Marge Ginsburg, Executive Director, Center for Healthcare Decisions
David Grossman, MD, MPH, Medical Director, Population Health
Christine S. Hunter, MD (Co- Chair), Chief Medical Officer, US Office of Personnel Management
Jeffery Kelman, MMSc, MD, Chief Medical Officer, United Stated Department of Health and Human Services
Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group
J. Brent Pawlecki, MD, MMM, Chief Health Officer, The Goodyear Tire & Rubber Company
Susan Reinhard, PhD, RN, Senior Vice President, AARP Public Policy Institute
Eric C. Schneider, MD, MSc, FACP (Co-chair), Senior Vice President, Policy and Research, The Commonwealth Fund
Marcus Thygeson, MD, MPH, Chief Health Officer, Blue Shield of California

HEDIS EXPERT CODING PANEL Members:
Glen Braden, MBA, CHCA, Attest Health Care Advisors, LLC
Denene Harper, RHIA, American Hospital Association
DeHandro Hayden, BS, American Medical Association
Patience Hoag, RHIT, CPHQ, CHCA, CCS, CCS-P, Health Services Advisory Group
Nelly Leon-Chisen, RHIA, American Hospital Association
Tammy Marshall, LVN, Aetna
Alec McLure, RHIA, CCS-P, Verisk Health
Michele Mouradian, RN, BSN, McKesson Health Solutions
Craig Thacker, RN, CIGNA HealthCare
Mary Jane F. Toomey, RN CPC, Aetna Better Health

Technical Measurement Advisory Panel Members:
Andy Amster, MSPH, Kaiser Permanente
Kathryn Coltin, MPH, Independent Consultant
Lekisha Daniel-Robinson, Centers for Medicare and Medicaid Services
Marissa Finn, MBA, Cigna HealthCare
Scott Fox, MS, MEd, Independence Blue Cross

Carlos Hernandez, CenCalHealth
Kelly Isom, MA, RN, Aetna
Harmon Jordan, ScD, RTI International
Ernest Moy, MD, MPH, Agency for Healthcare Research and Quality
Patrick Roohan, New York State Department of Health
Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC
Natan Szapiro, Independence Blue Cross

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2011

Ad.3 **Month and Year of most recent revision:** 07, 2015

Ad.4 **What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.
THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1800
Measure Title: Asthma Medication Ratio
Measure Steward: National Committee for Quality Assurance
Brief Description of Measure: The percentage of patients 5-64 years of age who were identified as having persistent asthma and had a ratio of controller medications to total asthma medications of 0.50 or greater during the measurement year.
Developer Rationale: This measure assesses patients with persistent asthma whose asthma is being controlled through the use of long-term asthma controller medications. The improvement in quality envisioned by the use of this measure is for plans to identify patients who are frequently using short-term asthma reliever medications to treat asthma exacerbations or acute symptoms and to increase their adherence to long-term controller medication or adjust medication. Increasing use of reliever medication or use more than two days a week for symptom control indicates the need to step up controller therapy (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007). According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered "not well controlled" or "very poorly controlled" as defined by clinical practice guidelines (Stillman 2010). Increasing use and adherence to asthma controller medications can prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami 2009; NHLBI/NAEPP 2007).

Akinbami, L.J., J.E. Moorman, P.L. Garbe, E.J. Sondik. 2009. Status of Childhood Asthma in the United States, 1980–2007. Pediatrics 123;S131-45. doi: 10.1542/peds.2008-2233C. (July 8, 2014).

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf (July 8, 2014).

Stillman, L. 2010. Living with Asthma in New England: Results from the 2006 BRFSS and Call-back Survey. A report by the Asthma Regional Council of New England (February). http://hria.org/resources/reports/asthma/living-with-asthma-in-new-england.html (July 8, 2014).

Numerator Statement: The number of patients who had a ratio of controller medications to total asthma medications of 0.50 or greater during the measurement year.

Denominator Statement: All patients 5-64 years of age as of December 31 of the measurement year who have persistent asthma by meeting at least one of the following criteria during both the measurement year and the year prior to the measurement year:
• At least one emergency department visit with asthma as the principal diagnosis
• At least one acute inpatient claim/encounter with asthma as the principal diagnosis
• At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits.
• At least four asthma medication dispensing events

Denominator Exclusions: Exclude patients who had any of the following diagnoses any time during the patient's history through the end of the measurement year (e.g., December 31):
-COPD
-Emphysema

-Obstructive Chronic Bronchitis
-Chronic Respiratory Conditions Due To Fumes/Vapors
-Cystic Fibrosis
-Acute Respiratory Failure

Exclude any patients who had no asthma medications (controller or reliever) dispensed during the measurement year.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy
Level of Analysis: Health Plan, Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2012 **Most Recent Endorsement Date:** Jul 31, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary or **Summary of prior review**

- The evidence for this measure is based on clinical practice guidelines for the diagnosis and management of asthma from the National Heart and Lung and Blood Institutes (NHLBI). Dated 2007, this is a strong recommendation, graded Category A and includes randomized controlled trials (RCT). The evidence presented supports the recommendation that: "long-term control medications (including ICSs, inhaled long-acting bronchodilators, leukotriene modifiers, cromolyn, theophylline, and immunomodulators) be taken daily on a long-term basis to achieve and maintain control of persistent asthma."
- The guidelines referenced 529 studies related to pharmacologic therapy for asthma, 12 studies showing the clinical effects of inhaled corticosteroids produced consistent benefits and high magnitude, and nine studies showing the usefulness of short-acting beta2-agonist (SABA).

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☒ **The developer provided updated evidence for this measure:**

Updates: The 2007 guidelines cited update those published in 2004.

Exception to evidence: N/A

Guidance from the Evidence Algorithm: 1→3 →4→5 (highest eligible rating is HIGH)

Questions for the Committee:

- *Has the developer presented adequate evidence to support the use of controller therapy to total asthma medication ratio of >=0.5, as being ideal or optimal?*
- *The guidelines have been updated, but the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee believe there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provide the following information:

- Asthma is one of the most prevalent chronic diseases. In 2010, 25.7 million Americans had asthma, including 7 million children, 15.6 million adults under 65 and 3.1 million adults 65 and older (Akinbami et al. 2012). Asthma is responsible for more than 3,000 deaths in the United States annually (American Lung Association 2014) and accounted for more than $50 billion spent on healthcare in the United States in 2007, an increase of almost $2 billion from 2002 (CDC 2011).
- The following data are extracted from HEDIS data and demonstrate variations in the numerator:

  Commercial Rate Ages 5-64 (HMO and PPO Combined)
  YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
  2012 | 77% | 6% | 24% | 69% | 74% | 77% | 80% | 83% | 91% | 6%
  (Commercial HMO: N=195 plans; mean size per plan=1,364; Commercial PPO: N=193 plans; mean size per plan=1,722)
  2013 | 79% | 6% | 30% | 73% | 76% | 80% | 82% | 85% | 100% | 6%
  (Commercial HMO: N=198 plans; mean size per plan=1,264; Commercial PPO: N=195 plans; mean size per plan=1,714;)
  2014 | 77% | 7% | 29% | 69% | 74% | 77% | 81% | 83% | 100% | 7%
  (Commercial PPO: N=192 plans; mean size per plan=1,162; Commercial PPO: N=194 plans; mean size per plan=1,657)

  Medicaid HMO Rate Ages 5-64
  YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
  2012 | 57% | 9% | 28% | 45% | 53% | 58% | 63% | 67% | 85% | 10% (N=122 plans; mean size per plan=2,012)
  2013 | 65% | 9% | 40% | 53% | 60% | 66% | 71% | 76% | 84% | 10% (N=147 plans; mean size per plan=1,888)
  2014 | 59% | 9% | 21% | 48% | 54% | 61% | 65% | 70% | 82% | 11% (N=162 plans; mean size per plan=1,986)

Disparities

- NCQA does not currently collect performance data stratified by race, ethnicity, or language; performance of commercial plans vs. Medicaid can be considered a proxy.
  - Mean performance of Medicaid plans is significantly lower as is performance at each of the percentiles reported above.
- This measure has not been stratified by race and ethnicity. However, the developer notes researchers have explored disparities in asthma outcomes and in utilization to health care services among children with asthma. Children of low income families experience more urgent care visits, hospitalizations and mortality due to asthma when compared to the general public (CDC 2009). This population experience poor outcomes due to barriers including access to care, access to medication, and insurance coverage.

Question for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

***1a. Evidence** to Support Measure Focus*

Comments:

**The guidelines have been updated, but the underlying evidence for the measure has not changed since the last NQF endorsement review. The ratio of >= 0.5 could be discussed further. Despite this, the evidence likely again meets an acceptable threshold.

**Measure #1800 is a process measure that is based on clinical practice guidelines by NHBLI that provides a Category A recommendation from RCT, with the evidence showing a direct correlation between use of long-term control medications taken daily for long periods of time for control and maintenance of asthma. The guidelines are based on an additional 550 studies that examine pharmacologic intervention of inhaled asthma medications, both long and short-acting.

**This is a process outcome that does have a direct relationship with intermediate outcome being measured. Patients with a ratio of 0.5 or greater experience fewer asthma exacerbations (as defined as either ED or acute inpatient visits) with asthma listed as the primary diagnosis.

This is specifically a process measure for asthma control and patients with asthma medication ration of 0.5 or greater indicate that they are using reliever medications less frequently to control the symptoms which points to the relationship between the measured outcome and the health outcome. This process measure is cited by the NHLBI/National Asthma Education and Prevention Program expert panel to support this measure.

1b. Performance Gap

Comments:

**A gap in care remains that warrants a national performance measure.

**Performance data was provided, showing a disparity between patients utilizing commercial insurance versus Medicaid, with Medicaid having a lower mean of the numerator. Asthma was also identified as one of the most prevalent chronic diseases with high incidence and high cost. Disparities by race, ethnicity, and language were not identified; however, some concessions were made for lower income families were noted (higher incidence of non-control of asthma).

**Performance data was measured through data extracted from HEDIS for 2012-2014. The performance data were summarized at the health plan level and stratified by year and product line (Medicaid, HMO, PPO). The data demonstrate the variation in the percentage of children and adults with persistent asthma and identify gaps in performance which indicates that these data may help to identify opportunities for improvement.

Data were stratifies by type of insurance, but data were not stratified by race, ethnicity, or language. These demographic data are difficult to collect at the plan level, but if they were able to be reasonably collected may provide additional information about disparities in care.

1c. High Priority (previously referred to as High Impact)

Comments:

**n/a

**N/A

**This measure does logically make a case for the overall quality construct, component performance measures, and their relationships. This measure specifically addresses a national health goal/priority identified by DHHS. This is a high priority area that impacts a large number of patients and poor asthma control has significant implications on morbidity and mortality, healthcare utilization, and societal consequences.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy

Specifications:

- The developer attests the specifications have not changed since the last submission.
    - The numerator of this measure is: *Number of patients who had a ratio of controller medications to total asthma medications of 0.50 or greater during the measurement year (12 month period).*
    - The denominator is: *Patients 5-64 years of age as of December 31 of the measurement year who have persistent asthma by meeting at least one of the following criteria during both the measurement year and the year prior to the measurement year:*
        - *At least one emergency department visit with asthma as the principal diagnosis*
        - *At least one acute inpatient claim/encounter with asthma as the principal diagnosis*
        - *At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits*
        - *At least four asthma medication dispensing events*
    - The ICD-9 and ICD-10 codes have been included in the specification details.
    - The calculation algorithm is stated in S.18 and appears straightforward.

Question for the Committee:

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Measure score reliability was assessed using data from health plans that submitted HEDIS data to NCQA in 2012 and 2015. In the newly submitted information, the developer provides the 2015 measure score reliability results, which used data for measurement year 2014 (386 commercial health plans and 164 Medicaid health plans).
- Past review noted concerns about clinical exclusions and medication inclusions.

Describe any updates to testing

- Reliability testing has been updated with more current data.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Reliability testing was conducted at the performance measure score level.
- Testing was conducted using signal to noise analysis.

Results of reliability testing

- The reliability statistics ranged from 0.93-0.97, which the developer states indicates strong reliability

Guidance from the Reliability Algorithm: 1 → 2 → 4 → 5 → 6/7 (highest eligible rating is HIGH)

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**

Questions for the Committee:

- *Are the specifications consistent with the evidence?*
- *Has the developer presented adequate evidence to support the use of controller therapy to total asthma medication ratio of >=0.5, as being ideal or optimal?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The prior Committee's vote was moderate; the Committee expressed concerns about the exclusions, particularly in the older age cohort.

Describe any updates to validity testing

- Additional empirical validity testing of the measure score has been conducted since the last review of this measure.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- The measure was tested for face validity with input from three expert panels.
- Empirical testing was conducted at the level of the performance measure score. Specifically, construct validity was assessed by examining whether the score for this measure was correlated with similar measures of respiratory care. The developer examined this measure against two other HEDIS measures:
  - Percent of patients who achieved asthma medication adherence of 75% or greater
  - Percent of patients who achieved asthma medication adherence of 50% or greater

Validity testing results:

The developer provides the following:

- Construct validity testing indicates that the asthma measures were significantly ($p<.05$) correlated with each other. Correlation coefficients were greater than 0.3 for both commercial and Medicaid plans, which the developer states ranges from moderate to strong.
- The expert panels concluded that the measures as specified will accurately differentiate quality across providers, thus having sufficient face validity.
- The measure performance rate was stratified by age group.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer notes several exclusions, as follows:
  - People with a diagnosis for a specific clinical condition (COPD, emphysema, obstructive chronic bronchitis, cystic fibrosis and acute respiratory failure)
  - Health plan members who do not meet one of four criterion (including asthma medication dispensing events and/or encounters with an asthma diagnosis) in both the measurement year and the year prior.
- 25% of commercial plan members and 18% of Medicaid plan members were excluded from the measure. A higher percentage of people ages 51-64 years had at least 1 measure exclusion compared to children or adults ages 5-50.

Questions for the Committee:

- *Are the results from the exclusion analysis a threat to validity?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Question for the Committee:

- *Should there be any risk adjustment? (Results are stratified by age group, per the developer.)*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer reports the following:

- Meaningful difference in performance is calculated using an inter-quartile range (IQR) for each indicator.
- Results indicate a 6-13% gap in performance between the 25th and 75th percentile performing plans. The highest variation in performance is for adults ages 19-50 and 51-64 in Medicaid plans.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- The developer states its audit process verifies that plans' measure calculations are not biased due to missing data.

Guidance from the Validity Algorithm: 1 → 2 → 3 → 6 → 7 → 8 (highest eligible rating is HIGH)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**Further discussion could be had whether the developer presented adequate evidence to support the use of controller therapy to total asthma medication ratio of >=0.5, as being ideal or optimal.

**Measure developer did not note why specifications may not be fully consistent with evidence.

**25% of commercial members and 18% of Medicaid members were excluded from the measure. The authors of the measure indicated that this measure is really designed to focus only on children and adults with persistent asthma and they anticipated that there would be more exclusions as the population aged secondary to a higher prevalence of COPD or chronic bronchitis. As the population ages and individuals develop additional diagnoses, it may be important to assess management and adherence in light of additional comorbities and stratify the data. It does appear that this is the major threat to the measure's validity.

2a2. Reliability Testing

Comments:

**Expert panels concluded that the measures as specified will accurately differentiate quality across providers, thus having sufficient face validity. Conclusions about quality can be made.

**Validity was noted as not being fully consistent with evidence, leading to at least a moderately low level score. However, validity testing level was at measure score with both Face and Empirical validity testing conducted, with correlation to similar respiratory care measures. P-value was low, demonstrating significant correlation and expert panel assessment showed that measures would differentiate quality across providers.

**The team tested for construct validity by exploring whether or not the measure was correlated with other similar measures of

respiratory care. The validity of performance was assessed by the Respiratory Measurement Advisory Panel and the Committee on Performance Management and it was determined that the asthma ratio measure was an accurate representation of quality performance and distinguished performance among health plans.

2b2. Validity Testing
Comments:
**The developer states its audit process verifies that plans' measure calculations are not biased due to missing data.
**No missing data was noted in the profile.
**NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Reliability testing was conducted at the performance measure score level.
**This appears to be a very reliable measure. The Beta-binomial model was used to estimate reliability. This measure suggests that this measure has good reliability. The vast majority of plans exceeded the minimally accepted threshold of 0.7 and the majority of plans exceeded 0.9.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer notes:

- All data elements are in defined fields in electronic claims.
- The data are generated during care processes.
- NCQA conducts an independent audit to verify HEDIS specifications are met and "real time" feedback is received from measure users.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**The developer reports no challenges or unexpected findings in implementation.
**All required data elements are included in electronic claims and are able to be verified via independent audit.
**This measure is feasible as all data elements are in defined fields in electronic claims.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure is publically reported nationally and by geographic regions in NCQA's State of Health Care annual report.
- The measure is reported in *Consumer Reports* and on the NCQA website, and is used to calculate NQCA's health plan ratings.
- The measure is used in scoring for accreditation of Medicare Advantage Health Plans
- This measure is used in Quality Compass.

Improvement results

The developer provides the following information:

- 2012-2014 – the measure showed slight improvement (approximately 2 percentage points) across Medicaid health plans
- Improvement has not been shown for commercial plans. There has been increasing attention to this measure in NCQA's health plan accreditation program; the developers hope this will help to improve the performance rates for commercial plans.

Unexpected findings (positive or negative) during implementation: The developer reports no challenges or unexpected findings in implementation.

Potential harms: The developer reports no unintended consequences were noted during testing.

Feedback: No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Although the measure is in use, is the small (to no) improvement over time indicative of poor usability?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Measure is in use and available. The small improvement differences across Medicaid plans over time and lack of improvement for commercial plans could be discussed further at the group level in terms of usability.
**The measure is currently being publicly reported in an accountability program, in NCQA's State of Health Care annual report, Consumer Reports, on the NCQA website, Quality Compass, and for accreditation of Medicare Advantage health plans. The measure can be utilized to target lower-performing populations as related to asthma control, but there may be concern that this measure with penalize those disparate areas with higher incidence rates and lower community infrastructure for quality care.
**The current measure is being publicly reported in several ways: Health Plan Rating, Annual State of Health Care Quality, Health Plan Accreditation and Quality Compass (links included in the measure document). This is a fairly straightforward process measure that can be used to track performance. I do not identify and unintended consequences that would make the risks of using this measure outweigh the benefits.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0047: Asthma: Pharmacologic Therapy for Persistent Asthma
- 1799: Medication Management for People with Asthma

Harmonization

Measures have not been harmonized. Developer provided the following rationale:

- 0047: Measure focus is different from #1800. Therefore, there is no impact on interpretability or added burden

of data collection. Both measures use different value sets to identify asthma controller medications.

Pre-meeting public and member comments

- None

NOTE: evidence that was included in the 2012 submission form may not completely match with the layout of the current evidence form.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1800

Measure Title: Asthma Medication Ratio

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.

- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: ratio of long-term asthma controller medications to short-term asthma reliever medications

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Members with a ratio of 0.5 or greater experience significantly fewer asthma exacerbations defined as either emergency department (ED) or acute inpatient visits with asthma listed as the primary diagnosis. The intent of the measure is to have members utilize both controllers and relievers in their regimens, instead of relievers alone thereby minimizing the number of preventable asthma exacerbations.

Diagram:

Provider identifies patients with persistent asthma >>> Provider dispenses both long-term asthma controller medication to patients to be used on a daily basis to manage and control asthma symptoms and inhaled short-acting beta2-agonist (SABA) to be used infrequently for quick relief of symptoms

>>> Provider monitors patients for control, educates patients on the importance of asthma controller medication adherence and assesses patient adherence to medication >>> Prevention and control of asthma symptoms, improvement in quality of life, and reduction in the frequency and severity of asthma exacerbations (desired outcome).

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051.
http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

National Heart Lung and Blood Institute/National Asthma and Education Prevention Program (NHLBI/NAEPP) Guidelines for the Diagnosis and Management of Asthma, 2007.

PERSISTENT ASTHMA

The Expert Panel recommends the following therapy for persistent asthma:

- Daily long-term control medication is recommended for patients who have persistent asthma. The long-term control medication should be ones with anti-inflammatory effects. Of the available medications, ICSs are the most effective single agents (Evidence Category A, page 216).
- Quick-relief medication must be available to all patients who have persistent asthma. The Expert Panel recommends that SABAs are the drug of choice for treating acute asthma symptoms and exacerbations and for preventing EIB (Evidence A, page 235). SABA should be taken as needed to relieve symptoms. The intensity of treatment will depend on the severity of the exacerbation. Increasing use of SABA or use more than 2 days a week for symptom control (not for preventing exercise-induced asthma) indicates the need to step up therapy. The Expert Panel does not recommend regularly scheduled, daily, long-term use of SABA (Evidence Category A, page 236).

Monitoring and follow-up is essential (Evidence Category B, page 277).

- When initiating therapy, monitor at 2- to 6-week intervals to ensure that asthma control is achieved (Evidence Category D).
- Regular follow-up contacts at 1- to 6-month intervals, depending on level of control, are recommended to ensure that control is maintained and the appropriate adjustments in therapy are made: step up if necessary or step down if possible. Consider 3-month intervals if a step down in therapy is anticipated (Evidence Category D).

LONG-TERM CONTROL MEDICATIONS

The Expert Panel recommends that long-term control medications (including ICSs, inhaled long-acting bronchodilators, leukotriene modifiers, cromolyn, theophylline, and immunomodulators) be taken daily on a long-term basis to achieve and maintain control of persistent asthma. The most effective long-term-control medications are those that attenuate the underlying inflammation that is characteristic of asthma (Evidence Category A, page 216).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

- When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel.
- Category A: Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.
- Category B: RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.
- Category D: Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

- When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong.
- Category C: Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

This measure is a process measure of asthma control. Patients with persistent asthma who were dispensed a ratio of asthma controller medications to total asthma medications (controller + reliever medications) of 0.50 or greater indicate that they are using reliever medications less frequently to control symptoms, and therefore their asthma control is being maintained.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

- When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel.
- Category A: Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.
- Category B: RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.
- Category D: Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

- When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong.
- Category C: Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1997-2006

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The Guidelines for the Diagnosis and Management of Asthma referenced a total of 1,654 studies to update the previous set of guidelines from 2004. The guidelines referenced 529 studies related to pharmacologic therapy for asthma, which included meta-analyses, systematic reviews of randomized controlled trials (RCTs), case control and cohort studies and non-analytic studies including case reports and case series. The guideline developers did not provide a breakdown of the specific number of randomized control trials (RCTs) and given the number of studies included in the systematic review we were not able to delineate all RCTs for each recommendation. However, the evidence review table related to inhaled corticosteroid dosing, for example, included 34 RCTs or meta-analyses and systematic reviews of RCTs. This review is not comprehensive and represents only a portion of the research on this area.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall, the quality of the evidence regarding daily long-term asthma controller medication for patients who have persistent asthma assessment is high. The 34 RCTs referenced above included thousands of patients studied over long periods of time. The evidence supporting the recommendation of daily long-term asthma controller medication for patients with persistent asthma was graded Category A, which includes randomized controlled trials (RCTs), a rich body of data, evidence from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made and requires substantial numbers of studies involving substantial numbers of participants.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The studies included evidence-based guidelines with and without systematic reviews/ evaluations, economic evaluations of asthma medications, survey based research and retrospective studies. Research and studies consistently show that appropriate medication management could potentially prevent a significant proportion of asthma-related costs.

The evidence for daily long-term asthma control medication in patients with persistent asthma shows consistent benefit and high magnitude. The guidelines referenced 12 studies showing that the clinical effects of inhaled

corticosteroids include reduction in severity of symptoms; improvement in asthma control and quality of life; improvement in peak expiratory flow and spirometry; diminished airway hyperresponsiveness; prevention of exacerbations; reduction in systemic corticosteroid courses, ED care, hospitalizations, and deaths due to asthma; and possibly the attenuation of loss of lung function in adults. Patients who have mild or moderate persistent asthma and are treated with ICS, compared to other single long-term control medications, demonstrate greater improvements in prebronchodilator forced expiratory volume (but not with long-term postbronchodilator forced expiratory volume); reduced airway hyperresponsiveness, symptom scores, exacerbation rates, and symptom frequency; as well as less use of supplemental SABA, fewer courses of oral systemic corticosteroids, and lower rates of hospitalization. The guideline also cites 9 studies showing that the frequency of SABA use can be clinically useful as a barometer of disease activity because increasing use of SABA has been associated with poor outcomes such as increased risk for death or near death in patients or increased risk of an acute exacerbation that requires an ED visit or hospitalization.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Benefits:

Prevention and management of asthma symptoms

Improved quality of life

Reduction in the frequency and severity of asthma exacerbations

Fewer ED visits

Harms: Potential adverse effects of long-term control and quick-relief medications

The majority of research on harms relates to potential side-effects of asthma medications. However, the guidelines state that "ICSs are the most effective long-term therapy available for mild, moderate, or severe persistent asthma; in general, ICSs are well tolerated and safe at the recommended

dosages (Evidence A). The potential but small risk of adverse events from the use of ICS treatment is well balanced by their efficacy."

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
AMR_Evidence-635899393674844305.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure assesses patients with persistent asthma whose asthma is being controlled through the use of long-term asthma controller medications. The improvement in quality envisioned by the use of this measure is for plans to identify patients who are frequently using short-term asthma reliever medications to treat asthma exacerbations or acute symptoms and to increase their adherence to long-term controller medication or adjust medication. Increasing use of reliever medication or use more than two days a week for symptom control indicates the need to step up controller therapy (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007). According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered "not well controlled" or "very poorly controlled" as defined by clinical practice guidelines (Stillman 2010). Increasing use and adherence to asthma controller medications can prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami 2009; NHLBI/NAEPP 2007).

Akinbami, L.J., J.E. Moorman, P.L. Garbe, E.J. Sondik. 2009. Status of Childhood Asthma in the United States, 1980–2007. Pediatrics 123;S131-45. doi: 10.1542/peds.2008-2233C. (July 8, 2014).

National Heart Lung and Blood Institute/National Asthma Education and Prevention Program. 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington (DC): National Heart Lung and Blood Institute (NHLBI), NIH Publication No. 07-4051. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf (July 8, 2014).

Stillman, L. 2010. Living with Asthma in New England: Results from the 2006 BRFSS and Call-back Survey. A report by the Asthma Regional Council of New England (February). http://hria.org/resources/reports/asthma/living-with-asthma-in-new-england.html (July 8, 2014).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year and product line (i.e. commercial, Medicaid, HMO and PPO).
The following data demonstrate the variation in the percentage of children and adults with persistent asthma who had a ratio of controller medications to total asthma medications of 0.50 or greater across health plans. In 2014 there was a 14 percentage point difference between commercial plans in the 10th percentile and commercial plans in the 90th percentile and 22 percentage point difference for Medicaid plans. These gaps in performance underscore the opportunity for improvement.

Commercial Rate Ages 5-64 (HMO and PPO Combined)

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 77% | 6% | 24% | 69% | 74% | 77% | 80% | 83% | 91% | 6%
2013 | 79% | 6% | 30% | 73% | 76% | 80% | 82% | 85% | 100% | 6%
2014 | 77% | 7% | 29% | 69% | 74% | 77% | 81% | 83% | 100% | 7%

Medicaid HMO Rate Ages 5-64
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 57% | 9% | 28% | 45% | 53% | 58% | 63% | 67% | 85% | 10%
2013 | 65% | 9% | 40% | 53% | 60% | 66% | 71% | 76% | 84% | 10%
2014 | 59% | 9% | 21% | 48% | 54% | 61% | 65% | 70% | 82% | 11%

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Commercial HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 195 | 1,364
2013 | 198 | 1,264
2014 | 192 | 1,162

Commercial PPO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 193 | 1,722
2013 | 195 | 1,714
2014 | 194 | 1,657

Medicaid HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 122 | 2,012
2013 | 147 | 1,888
2014 | 162 | 1,986

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
HEDIS data are stratified by type of insurance (e.g. Commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.
Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts. Health Affairs 20(10): 1984-1991.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Although HEDIS measures are not stratified by race and ethnicity, researchers have explored disparities in asthma outcomes and in utilization to health care services among children with asthma. Children of low-income families experience more urgent care visits, hospitalizations and mortality due to asthma when compared to the general public (CDC 2009). Poor asthma outcomes in low-income children may be partly due to the barriers they face in accessing care for asthma, including access to medications. One study found that children with asthma from low-income families were less likely to have prescriptions filled and/or receive annual primary health examinations (Kim et al. 2009). The study also examined insurance coverage, showing that children without insurance coverage utilized primary health care services for asthma less often (Kim et al. 2009).

Centers for Disease Control and Prevention (CDC). Asthma: A Presentation of Asthma Management and Prevention, September 2009. http://www.cdc.gov/asthma/speakit/default.htm (July 8, 2014).

Kim, H., G.M. Kieckhefer, A.A. Greek, J.M. Joesch, N. Baydar. 2009. Health Care Utilization by Children With Asthma. Preventing Chronic Disease 6(1): A12.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Asthma is one of the most prevalent chronic diseases. In 2010, 25.7 million Americans had asthma, including 7 million children, 15.6 million adults under 65 and 3.1 million adults 65 and older (Akinbami et al. 2012). Asthma has also become increasingly more common over the past decade, occurring in 7.3 percent of the population in 2001 compared to 8.4 percent in 2010 (Akinbami et al. 2012). Asthma is responsible for over 3,000 deaths in the U.S. annually (American Lung Association 2014) and accounted for over $50 billion spent on health care in the United States in 2007, an increase of almost $2 billion from 2002 (CDC 2011).

Appropriate medication adherence could ameliorate the severity of many asthma-related symptoms (Akinbami et al. 2009). According to the Asthma Regional Council of New England, two-thirds of adults and children who display asthma symptoms are considered “not well controlled” or “very poorly controlled” as defined by clinical practice guidelines (Stillman 2010). Pharmacologic therapy is used to prevent and control asthma symptoms, improve quality of life, reduce the frequency and severity of asthma exacerbations, and reverse airflow obstruction (National Heart, Lung, and Blood Institute [NHLBI]/National Asthma and Education Prevention Program [NAEPP] 2007). Appropriate medication management could potentially prevent a significant proportion of asthma-related costs (hospitalizations, emergency room visits and missed work and school days) (Akinbami et al. 2009). Indeed, several studies have found that higher medication adherence rates are associated with better outcomes; for instance, one study found that patients with asthma controller medication adherence rates of 75% or greater had fewer asthma exacerbations compared to patients with adherence rates of 25% or lower (Williams et al. 2011). The Asthma Regional Council has also stated that proper management could potentially save at least 25 percent of total asthma costs, or $5 billion, nationally by reducing health care costs (American Lung Association 2012).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Akinbami, L.J., J.E. Moorman, P.L. Garbe, E.J. Sondik. 2009. Status of Childhood Asthma in the United States, 1980–2007. Pediatrics 123;S131-45. doi: 10.1542/peds.2008-2233C.

Akinbami, L.J., J.E. Moorman, C. Bailey, H.S. Zahran, M. King, C.A. Johnson, X. Liu. 2012. “Trends in Asthma Prevalence, Health Care Use, and Mortality in the United States, 2001-2010.” NCHS Data Brief, no. 94 (May). http://www.cdc.gov/nchs/data/databriefs/db94.pdf (November 19, 2015).

American Lung Association. 2012. Trends in Asthma Morbidity and Mortality. http://www.lung.org/finding-cures/our-research/trend-reports/asthma-trend-report.pdf (November 19, 2015).

American Lung Association. 2014. Asthma & Children Fact Sheet, September. http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/asthma/learn-about-asthma/asthma-children-facts-sheet.html (November 19, 2015).

Centers for Disease Control and Prevention (CDC). Vital Signs: Asthma in the US, May 2011. http://www.cdc.gov/VitalSigns/Asthma/index.html (November 19, 2015).

Stillman, L. 2010. Living with Asthma in New England: Results from the 2006 BRFSS and Call-back Survey. A report by the Asthma Regional Council of New England (February). http://hria.org/resources/reports/asthma/living-with-asthma-in-new-england.html (November 19, 2015).

Williams, L. K., E.L. Peterson, K. Wells, B.K. Ahmedani, R. Kumar, E.G. Buchard, V.K. Chowdhry, D. Favro, D.E. Lanfear, M. Pladevall. 2011. Quantifying the proportion of severe asthma exacerbations attributable to inhaled corticosteroid nonadherence. Journal of Allergy and Clinical Immunology, no 128.6 p. 1185-1191

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 1800_AMR_Value_Sets.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no significant changes to the measure specification since the last endorsement maintenance completed on January 31, 2012.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of patients who had a ratio of controller medications to total asthma medications of 0.50 or greater during the measurement year.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: 12 month period (the measurement year)
Denominator: 24 month period (the measurement year and the year prior)
Exclusions: lookback through the patient's history through the last day of the measurement year

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Follow the steps below to identify numerator compliance.

Step 1: For each patient, count the units of controller medications (see AMR-A) dispensed during the measurement year. When identifying medication units for the numerator, count each individual medication, defined as an amount lasting 30 days or less, as one medication unit. One medication unit equals one inhaler canister, one injection, or a 30-day or less supply of an oral medication. For example, two inhaler canisters of the same medication dispensed on the same day count as two medication units and only one dispensing event. Use the package size and units columns in the NDC list to determine the number of canisters or injections. Divide the dispensed amount by the package size to determine the number of canisters or injections dispensed. For example, if the package size for an inhaled medication is 10g and pharmacy data indicates the dispensed amount is 30 g, this indicates 3 inhaler canisters were dispensed.

Step 2: For each patient, count the units of reliever medications (see AMR-A) dispensed during the measurement year.

Step 3: For each patient, sum the units calculated in step 1 and step 2 to determine units of total asthma medications.

Step 4: For each patient, calculate the ratio of controller medications to total asthma medications using the following formula: Units of Controller Medications (Step 1)/ Units of Total Asthma Medications (Step 3)

Step 5: Sum the total number of patients who have a ratio of 0.50 or greater in step 4.

AMR-A: Asthma Controller and Reliever Medications
Asthma Controller Medications:
-Antiasthmatic combinations: dyphylline-guaifenesin; guaifenesin-theophylline

-Antibody inhibitors: omalizumab

-Inhaled steroid combinations: budesonide-formoterol; fluticasone-salmeterol; mometasone-formoterol

-Inhaled corticosteroids: beclomethasone; budesonide; ciclesonide; flunisolide; fluticasone CFC free; mometasone

-Leukotriene modifiers: montelukast; zafirlukast; zileuton

-Mast cell stabilizers: cromolyn

-Methylxanthines: aminophylline; dyphylline; theophylline.

Asthma Reliever Medications:
-Short-acting, inhaled beta-2 Agonists: albuterol; levalbuterol; pirbuterol.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients 5–64 years of age as of December 31 of the measurement year who have persistent asthma by meeting at least one of the following criteria during both the measurement year and the year prior to the measurement year:
• At least one emergency department visit with asthma as the principal diagnosis
• At least one acute inpatient claim/encounter with asthma as the principal diagnosis
• At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits.
• At least four asthma medication dispensing events

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Individuals with multiple chronic conditions

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The eligible population for the denominator is defined by following the series of steps below:

Step 1: Identify patients as having persistent asthma who met at least one of the following criteria during both the measurement year and the year prior to the measurement year. Criteria need not be the same across both years.
• At least one ED visit (refer to codes in ED Value Set) with asthma as the principal diagnosis (refer to codes in Asthma Value Set).
• At least one acute inpatient claim/encounter (refer to codes in Acute Inpatient Value Set) with asthma as the principal diagnosis (refer to codes in Asthma Value Set).
• At least four outpatient visits (refer to codes in Outpatient Value Set) or observation visits (refer to codes in Observation Value Set) on different dates of service, with any diagnosis of asthma (refer to codes in Asthma Value Set) AND at least two asthma medication dispensing events (see MMA-A). Visit type need not be the same for the four visits.
• At least four asthma medication dispensing events (see MMA-A)

Step 2: A patient identified as having persistent asthma because of at least four asthma medication dispensing events, where leukotriene modifiers or antibody inhibitors were the sole asthma medication dispensed in that year, must also have at least one diagnosis of asthma (refer to codes in Asthma Value Set), in any setting, in the same year as the leukotriene modifier or antibody inhibitor (i.e., measurement year or year prior to the measurement year).

See attached value set Excel document for the following value sets:
- ED Value Set
- Asthma Value Set
- Acute Inpatient Value Set
- Outpatient Value Set
- Observation Value Set

MMA-A: Asthma Medications
Antiasthmatic combinations: dyphylline-guaifenesin; guaifenesin-theophylline

Antibody inhibitor: omalizumab

Inhaled steroid combinations: budesonide-formoterol; fluticasone-salmeterol; Mometasone-formoterol

Inhaled corticosteroids: beclomethasone; budesonide; ciclesonide; flunisolide; fluticasone CFC free; mometasone

Leukotriene modifiers: montelukast; zafirlukast; zileuton

Mast cell stabilizers: cromolyn

Methylxanthines: aminophylline; dyphylline; theophylline

Short-acting, inhaled beta-2 Agonists: albuterol; levalbuterol; metaproterenol; pirbuterol

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

Exclude patients who had any of the following diagnoses any time during the patient's history through the end of the measurement year (e.g., December 31):
-COPD
-Emphysema
-Obstructive Chronic Bronchitis
-Chronic Respiratory Conditions Due To Fumes/Vapors
-Cystic Fibrosis
-Acute Respiratory Failure

Exclude any patients who had no asthma medications (controller or reliever) dispensed during the measurement year.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1) Exclude patients who had any diagnosis of Emphysema (refer to codes in Emphysema Value Set or Other Emphysema Value Set), COPD (refer to codes in COPD Value Set), Chronic Bronchitis (refer to codes in Obstructive Chronic Bronchitis Value Set), Chronic Respiratory Conditions Due To Fumes/Vapors (refer to codes in Chronic Respiratory Conditions Due to Fumes/Vapors Value Set), Cystic Fibrosis (refer to codes in Cystic Fibrosis Value Set) or Acute Respiratory Failure (refer to codes in Acute Respiratory Failure Value Set) any time during the patient's history through the end of the measurement year (e.g., December 31).

2) Exclude any patients who had no asthma medications (controller or reliever) (see AMR-A) dispensed during the measurement year.

See attached value set Excel document for the following value sets:
- Emphysema Value Set
– Other Emphysema Value Set
– COPD Value Set
– Obstructive Chronic Bronchitis Value Set
– Chronic Respiratory Conditions Due to Fumes/Vapors Value Set
– Cystic Fibrosis Value Set
– Acute Respiratory Failure Value Set

AMR-A: Asthma Controller and Reliever Medications:
Asthma Controller Medications:
Antiasthmatic combinations: dyphylline-guaifenesin; guaifenesin-theophylline

Antibody inhibitors: omalizumab

Inhaled steroid combinations: budesonide-formoterol; fluticasone-salmeterol; mometasone-formoterol

Inhaled corticosteroids: beclomethasone; budesonide; ciclesonide; flunisolide; fluticasone CFC free; mometasone;

Leukotriene modifiers: montelukast; zafirlukast; zileuton

Mast cell stabilizers: cromolyn

Methylxanthines: aminophylline; dyphylline; theophylline.

Asthma Reliever Medications:
Short-acting, inhaled beta-2 Agonists: albuterol; levalbuterol; pirbuterol.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Four age stratifications and a total rate are reported for this measure. Age for each stratum is based on the patient's age as of the end of the Measurement Year (e.g., December 31).
1) 5–11 years

2) 12–18 years
3) 19-50 years
4) 51-64 years
5) Total (5-64 years)

The age strata align with both clinical practice guidelines and reporting requirements for child health quality improvement programs. Clinical guidelines specify appropriate age cohorts for measuring use of asthma medications as 5–11 years of age and 12–50 years of age, to account for the differences in medication regimens for children compared to adolescents and adults. Implementation requires further stratification of the age ranges to enable creation of comparable cohorts that align with child health populations.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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.)*
Refer to items S.6 (Numerator details), S.9 (Denominator details), S.11 (Denominator exclusions details) and S.2b (Data Dictionary) for tables.

This measure determines the percentage of patients with persistent asthma who had a ratio of controller medications to total asthma medications of 0.50 or greater based on information available from the published NDC codes. The measure calculation is detailed in the steps listed below:

Step 1: Determine the eligible population: Identify patients 5–64 years of age as of December 31 of the measurement year as having persistent asthma who met at least one of the following criteria during both the measurement year and the year prior to the measurement year. Criteria need not be the same across both year:
a) At least one ED visit with asthma as the principal diagnosis; or
b) At least one acute inpatient claim/encounter with asthma as the principal diagnosis; or
c) At least four outpatient visits or observation visits on different dates of service, with any diagnosis of asthma AND at least two asthma medication dispensing events. Visit type need not be the same for the four visits; or
d) At least four asthma medication dispensing events*

*A patient identified as having persistent asthma because of at least four asthma medication dispensing events where leukotriene

modifiers or antibody inhibitors were the sole asthma medication dispensed in that year, must also have at least one diagnosis of asthma, in any setting, in the same year as the leukotriene modifier or antibody inhibitor (i.e., measurement year or year prior to the measurement year).

Step 2: Determine denominator exclusions:
a) Exclude patients who had any diagnosis of Emphysema, COPD, Chronic Bronchitis, Chronic Respiratory Conditions Due to Fumes/Vapors, Cystic Fibrosis or Acute Respiratory Failure any time during the patient's history through the end of the measurement year
b) Exclude patients who had no asthma medications (controller or reliever) dispensed during the measurement year.

Step 3: Determine numerator:
a) For each patient, count the units of controller medications (see AMR-A) dispensed during the measurement year. When identifying medication units for the numerator, count each individual medication, defined as an amount lasting 30 days or less, as one medication unit. One medication unit equals one inhaler canister, one injection, or a 30-day or less supply of an oral medication. For example, two inhaler canisters of the same medication dispensed on the same day count as two medication units and only one dispensing event. Use the package size and units columns in the NDC list to determine the number of canisters or injections. Divide the dispensed amount by the package size to determine the number of canisters or injections dispensed. For example, if the package size for an inhaled medication is 10g and pharmacy data indicates the dispensed amount is 30 g, this indicates 3 inhaler canisters were dispensed.
b) For each patient, count the units of reliever medications (see AMR-A) dispensed during the measurement year.
c) For each patient, sum the units calculated in step a and step b to determine units of total asthma medications.
d) For each patient, calculate the ratio of controller medications to total asthma medications using the following formula:
Units of Controller Medications (Step a)/ Units of Total Asthma Medications (Step c)
e) Sum the total number of patients who have a ratio of 0.50 or greater in step d.

Step 4: Calculate the measure rate: the number of patients have a ratio of 0.50 or greater/Denominator

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at*

A.1)
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
AMR_Testing-635899393905895115.docx

NOTE: testing information that was included in the 2012 submission form may not completely match with the layout of the current testing form.

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**: 1800**
Measure Title: **Asthma Medication Ratio**
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in

understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items.

Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
N/A

1.3. What are the dates of the data used in testing? Click here to enter date range
Initial testing: During measure development, we conducted a comprehensive field test to assess feasibility of data collection and validity of the numerator, denominator and exclusions. This field test used data from measurement year 2009, which included health plan data spanning January 1, 2008 through December 31, 2009.

Systematic evaluation of face validity: The measure was tested for face validity throughout measure development from 2010 to 2012. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2012.

2015 Update: We assessed measure score reliability and construct validity using data from all health plans that submitted HEDIS data to NCQA for this measure in 2015, which used data for measurement year 2014. Measurement year 2014 required two years' worth of health plan data from January 1, 2013 through December 31, 2014.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Initial testing: During measure development, we conducted a comprehensive field test to assess feasibility of data collection and validity of the numerator, denominator and exclusions. This field test used data from measurement year 2009, which included health plan data spanning January 1, 2008 through December 31, 2009.

Systematic evaluation of face validity: Throughout the entire measure development process from 2010-2012, the measure was tested for face validity using panels of experts with specific clinical, methodologic and operational expertise. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2012. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliation of expert panel:

1) NCQA's Respiratory Measurement Advisory Panel (RMAP) is comprised of 10 experts (8 physicians, 1 pharmacist and 1 researcher) in clinical pulmonary care, including health care providers and policy makers.

2) NCQA's Technical Measurement Advisory Panel is a 12-member panel representing health plans methodologists, clinicians and HEDIS auditors.

3) NCQA's Committee on Performance Measurement (CPM) oversees the HEDIS measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 17 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

4) [2015 Update] NCQA's HEDIS Expert Coding Panel reviewed and provided feedback on the vocabularies and definitions found in the values sets used to identify each measure component as well as the more recent mapping of ICD-9 codes to ICD-10 codes.

In 2005, the draft measure was posted for public comment, a 30-day period of review that allowed interested parties to offer feedback to NCQA about the measure. Stakeholders from various types of organizations submitted 42 comments on the measure.

2015 Update: Measure score reliability and construct validity was calculated from the 386 Commercial health plans and 164 Medicaid health plans that submitted data on this measure to HEDIS in 2015. The plans were geographically diverse and varied in size.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Patient sample for initial measure field testing: We collected data from 9 plans (7 Commercial and 7 Medicaid plans). Below is a description of the sample. It includes the number of health plans that provided data for the measurement year 2009 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial	7	3,920
Medicaid	7	2,577

2015 Update: Measure score reliability and construct validity testing: In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the number of health plans that submitted data for this measure to HEDIS for measurement year 2014 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial	386	566
Medicaid	164	1,123

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

During measure development, we conducted a comprehensive field test to assess feasibility of data collection, reliability and validity of the numerator, denominator and exclusions using data submitted by 9 plans (7 Commercial and 7 Medicaid plans). The field test used the measurement year 2009 that included data from January 1, 2008 through December 31, 2009.

Face validity was demonstrated through a systematic assessment of face validity during measure development and at regular intervals since then. Per NQF instructions we have described the composition of the technical expert panel which assessed face validity in the data sample questions above.

2015 Update: The measure underwent additional analyses to assess measure score reliability (tested using a beta-binomial calculation) and construct validity (tested using Pearson's correlations of similar measures). These analyses included all of the health plans (386 Commercial and 164 Medicaid) that submitted data for this measure to HEDIS for measurement year 2014 (using data from January 1, 2013 through December 31, 2014).

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Measure performance results are stratified by commercial and Medicaid health plans and by age (5-11 years; 12-18 years; 19-50 years; 51-64 years).

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e*.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Beta-Binomial Statistic:

Commercial			Medicaid		
Median	Overall	10th-90th	Median	Overall	10th-90th
0.93	0.87	0.66-0.99	0.97	0.93	0.81-0.99

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Interpretation of measure score reliability testing:
Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. Asthma beta binomial testing suggests that this measure has good reliability. The 10-90th percentile distribution of health plan level-reliability on this measure show the vast majority of health plans exceeded the minimally accepted threshold of 0.7, and the majority of plans exceeded 0.9. Strong reliability is demonstrated since majority of variances is due to signal and not to noise.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

METHOD OF ASSESSING FACE VALIDITY

2012 Submission Form [Testing Data]: NCQA tested the measure for face validity using a panel of stakeholders with relevant clinical expertise and research and measurement, experience. This panel included representatives from key stakeholder groups, including the CDC, pulmonologists, provider and deliver organizations and researchers (See list of members for the Respiratory Advisory Panel (RMAP) under section Ad.1). RMAP experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2015 Update: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

The asthma medication ratio measure was developed in 2010 to assess patients with persistent asthma whose asthma is being controlled through long-term asthma controller medications. NCQA and the Respiratory Measurement Advisory Panel worked together to develop the most appropriate measure for assessing asthma medication adherence.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

The asthma medication ratio measure was written and field-tested in 2010. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in 2011.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

The asthma medication ratio measure was released for Public Comment in 2011 prior to publication in HEDIS. We received and responded to 42 comments on this measure. The CPM recommended moving this measure to first year data collection by a majority vote.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

The asthma medication ratio measure was introduced to HEDIS in 2011. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. Slight adjustments were made to the measure and it was again analyzed for public reporting in 2013. The CPM recommended moving this measure public reporting with a majority vote in 2013.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publically reported and may be used for scoring in accreditation.

The asthma medication ratio measure has been publicly reported in HEDIS since 2013.

STEP 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

The clinical guideline recommendations for medication management of asthma have not changed since the measure was developed in 2010; therefore, we have not made any significant changes to the medication ratio measure since it was last endorsed on January 31, 2012.

Expert Participation

This measure was tested for face validity with input from three expert panels. Guidelines from the National Heart, Lung and Blood Institute/National Asthma Education and Prevention Program in 2007 were also a strong authoritative source in applying the evidence for the asthma medication ratio measure.

We list an overview of each panel here. Please refer to Ad.1 in the submission form for the names and affiliation of experts in each panel.

1) Respiratory Measurement Advisory Panel includes 10 members (eight physicians, one pharmacist and a researcher) with expertise in respiratory care and quality measurement.

2) The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.

3) NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

ICD-10 CONVERSION:

In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA first identified value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM)to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.
3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented to for expert review and feedback.
4. NCQA RMAP clinical review: Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.

5. New value sets containing ICD-10 code recommendations were for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under **Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development**.

METHOD OF TESTING CONSTRUCT VALIDITY

2015 Update: We tested for construct validity by exploring whether this measure was correlated with other similar measures of respiratory care. We hypothesized that organizations that perform well on the measure should perform well on other similar HEDIS measures. To test these correlations we used a Person correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

For this measure, we specifically hypothesized:
1) The asthma medication ratio measure will be positively correlated with the medication management for people with asthma indicator (percent of patients who achieved asthma medication adherence of 75% or greater)
2) The asthma medication ratio measure will be positively correlated with the medication management for people with asthma indicator (percent of patients who achieved asthma medication adherence of 50% or greater)

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

RESULTS OF FACE VALIDITY ASSESSMENT

2012 Submission Form [Testing Data]:
For the initial field test in 2010, we calculated the measure performance rate stratified by age group and discussed the validity of the performance results with the expert panels (Respiratory Measurement Advisory Panel and the Committee on Performance Measurement). The expert panels agreed that the performance on the asthma ratio measure was an accurate representation of quality performance and distinguished performance among health plans.

2010 Field Test: Performance Rates on the Asthma Medication Ratio Measure*

	Age Group	Denominator	Numerator	Performance Rate
Commercial	**Ages 5-11**	5,670	3,454	**60.9%**
	Ages 12-50	19,242	11,272	**58.6%**
	Ages 51-64	10,944	7,352	**67.2%**
	Total (Ages 5-64)	35,856	22,078	**61.6%**
Medicaid	**Ages 5-11**	8,301	4,934	**59.4%**
	Ages 12-50	11,794	5,540	**47.0%**
	Ages 51-64	1,529	740	**48.4%**
	Total (Ages 5-64)	21,624	11,214	**51.9%**

**Includes data submitted by 7 Commercial plans and 7 Medicaid plans using measurement year 2009*

We assessed validity of the denominator criteria to ensure the measure is capturing people with persistent asthma. Entry into the eligible population for persistent asthma requires a combination of multiple outpatient encounters and

diagnoses. Approximately 90% of commercial members and 88% of Medicaid members were included in the eligible population by having at least four asthma medication dispensing events in the measurement year and the year prior. The remaining 10-12% of members had at least one ED visit with a diagnosis of asthma, one inpatient visit with a diagnosis of asthma, or four outpatient visits with a diagnoses of asthma plus two asthma medication dispensing events in the measurement year and the year prior.

We examined whether encounters could be linked to the same event and therefore do not accurately capture a population with persistent asthma. Using the field test dataset, NCQA examined the different scenarios where encounters were less than 14 days apart (a standard HEDIS time frame for linked encounters) to determine the effect on the measure's eligible population. Section 2b3.3 details the results of this additional analysis revealing the proportion of the population that would potentially excluded from the EP as a result of the additional criterion of <14 days between encounters. The next table details the proportion of the population that would potentially be excluded from the EP as a result of the additional criterion of <14 days between encounters. For this table, "Eligible Population" (EP) refers only to those members who satisfied the "Combination" eligibility criterion of at least four outpatient encounters and at least two prescription events.

Proportion of the Eligible Population affected by a ≥14 Day rule.

	Commercial				Medicaid			
		Do Not Qualify				Do Not Qualify		
Age Group	EP	N	% of EP	% of total EP	EP	N	% of EP	% of total EP
5-11	293	77	26.3%	1.3%	78	17	21.8%	0.2%
12-50	634	137	21.6%	0.6%	88	16	18.2%	0.1%
51-64	715	104	14.5%	0.6%	14	0	0.0%	0.0%
Total 1 (5-50)	927	214	23.1%	0.7%	166	33	19.9%	0.1%
Total 2 (5-64)	1,854	428	23.1%	0.9%	180	33	18.3%	0.1%

Another concern when measuring management for plan-to-plan comparison is ensuring that the majority of index prescriptions occur at a point within the measurement year (Q1, Q2, Q3, & Q4) that objectively monitors adherence without any type of adjustment. It addresses the question: Is the prescription utilization stable for this population and, if so, is the administrative data capturing index prescription start dates (IPSDs) sufficiently early in the measurement year to adequately measure medication management. The following table outlines the percentage of index prescriptions occurring in each quarter of the measurement year by cohort. The table presents the percentage of index prescriptions dispensed to members of the entire Eligible Population after comorbidity exclusions have been applied.

Timing of Index Prescription (by Age Group and line of business)

Product	Age	Q1	Q2	Q3	Q4
Commercial	5-11	67.2%	16.1%	6.9%	5.3%
	12-50	65.1%	14.3%	6.1%	4.3%
	51-64	72.7%	12.7%	4.0%	2.5%
Medicaid	5-11	62.7%	15.0%	6.1%	4.9%
	12-50	55.5%	12.9%	5.8%	2.1%
	51-64	58.9%	6.3%	2.8%	2.1%

The expert panels agreed that the denominator as specified was valid in identifying people with persistent asthma.

RESULTS OF CONSTRUCT VALIDITY TESTING

2015 Submission Form: The results indicated that the asthma measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized. The level of correlations among 2 out of the 3 indicators for both Commercial and Medicaid plans ranged from moderate to strong (the correlation coefficients were higher than 0.3).

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Commercial Plans)*

Asthma Measure	Pearson Correlation Coefficients		
	Medication Adherence 50%	Medication Adherence 75%	Asthma Medication Ratio
Medication Adherence 50%	-	0.9	0.3
Medication Adherence 75%	-	-	0.2

**Includes data submitted by 388 Commercial plans to HEDIS for these measures for measurement year 2014*

Note: All correlations are significant at p<.05

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Medicaid Plans)*

Asthma Measure	Pearson Correlation Coefficients		
	Medication Adherence 50%	Medication Adherence 75%	Asthma Medication Ratio
Medication Adherence 50%	-	1.0	0.3
Medication Adherence 75%	-	-	0.2

**Includes data submitted by 192 Medicaid plans to HEDIS for these measures for measurement year 2014*

Note: All correlations are significant at p<.05

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

SYSTEMATIC ASSESSMENT OF FACE VALIDITY

2015 Update: The current asthma medication ratio measure for children and adults ages 5-64 was deemed to have the desirable attributes of a HEDIS measure in 2011 (relevance, scientific soundness, and feasibility). The technical expert panels showed good agreement that the measure as specified accurately identifies patients with persistent asthma and differentiates quality across providers. Our interpretation of these results is that this measure has sufficient face validity.

CONSTRUCT VALIDITY

2015 Update: Coefficients with absolute value of less than 0.2 are generally considered indicative of weak associations whereas absolute values of 0.2 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that the asthma measures are correlated with each other, suggesting they represent the same underlying quality construct of asthma quality of care.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2012 Submission Form [Testing Data]: The measure is intended to assess patients with persistent asthma whose asthma is being controlled through long-term asthma controller medications and to align with the clinical guideline recommendations for medication management of persistent asthma. The measure is not intended to capture people with intermittent or seasonal asthma or those who have non-asthma respiratory conditions. To ensure the measure captures people with persistent asthma only, members must meet one of four criterion (including asthma medication dispensing events and/or encounters with an asthma diagnosis) in both the measurement year and the year prior. The measure also excludes people with a diagnosis for a specific clinical condition (COPD, emphysema, obstructive chronic bronchitis, cystic fibrosis and acute respiratory failure). During measure development in 2010, exclusions were tested

using data from 7 commercial and 7 Medicaid health plans to determine the impact each clinical condition had on the measure's performance. We calculated the percent of people excluded from the denominator (i.e., the percent who had at least 1 exclusion condition) and the total percent of people excluded from the denominator for each age cohort.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2012 Submission Form [Testing Data]: A total of 25% of Commercial members and 18% of Medicaid members were excluded from the measure. A higher percentage of people ages 51-64 had at least 1 measure exclusion compared to children or adults ages 5-50.

Impact of Co-morbidity Exclusions on the Eligible Population*

	Age Range	EP	Percent of Eligible Population Excluded for a Co-Morbidity					
			At least 1 exclusion	COPD	Chronic Bronchitis	Emphysema	Cystic Fibrosis	Acute Respiratory Syndrome
Comm. N=7 plans	5 - 11	6,031	5.7%	3.7%	1.3%	0.3%	0.8%	1.0%
	12 - 50	22,855	16.2%	14.2%	4.1%	0.9%	0.6%	1.3%
	51 - 64	18,154	41.5%	39.6%	15.6%	6.4%	0.1%	3.5%
	Total (5-64 years)	47,040	24.6%	22.6%	8.2%	3.0%	0.4%	2.1%
Medicaid N=7 plans	5 - 11	8,614	3.8%	2.7%	0.4%	0.1%	0.4%	0.6%
	12 - 50	14,337	18.3%	16.6%	3.4%	1.3%	0.4%	2.6%
	51 - 64	4,432	45.2%	43.7%	13.8%	6.7%	0.2%	7.4%
	Total (5-64 years)	27,383	18.1%	16.6%	4.2%	1.8%	0.3%	2.7%

EP= total eligible population (number of patients) across all health plans prior to exclusions
**Includes data submitted by 7 Commercial plans and 7 Medicaid plans using measurement year 2009*

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2015 Update: The denominator criteria and the exclusions in this measure are intended to focus the measure on children and adults who have persistent asthma rather than concomitant diagnoses of asthma and COPD or chronic bronchitis. The higher percentage of adults ages 51-64 being excluded from the measure was expected, as they have a higher prevalence of conditions such as COPD or chronic bronchitis. The measure exclusions are needed in order to: 1) optimize the specificity of the denominator to only include those with persistent asthma and to keep the measure aligned with the clinical guideline recommendations; and to 2) display performance results that truly reflect appropriate care for this cohort of patients.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2012 Submission Form [Testing Data]: Nine health plans covering a variety of geographic areas within the United States were asked to provide complete administrative data file consisting of any member in their commercial and Medicaid product lines for anyone that had a diagnosis code for asthma during the calendar years of 2009-2010. The complete member-level administrative file used for analysis included a total of more than 82,000 health plan members with

persistent asthma. Specific calculations involve average performance rate, distribution (percentiles), 95% confidence interval of average rate across the respective health plans per by product line.

2015 Update: To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measured entities. However the method can be used for comparison of any two measured entities.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? *(e.g., number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined)*

2012 Submission Form [Testing Data]: Distribution of plan performance for the field test data set by each product line (commercial and Medicaid).

Product Line	Ave Rate	Lower 95% CI	Upper 95% CI	Standard Deviation	Min Rate	Max Rate	10th	25th	50th	75th	90th
Comm.	0.66072	0.55895	0.76249	0.11004	0.48263	0.75188	0.48263	0.58115	0.73476	0.75098	0.75188
Medicaid	0.46997	0.37192	0.56803	0.10602	0.24638	0.57154	0.24638	0.46164	0.49164	0.53391	0.57154

2015 Update: HEDIS 2014 Variation in Performance across Health Plans, Asthma Medication Ratio*

	Ages	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Comm.	Ages 5-11	264	87	7	79	84	88	91	93	7	<0.001
	Ages 12-18	235	76	8	67	72	77	81	84	9	0.003
	Ages 19-50	642	71	8	62	66	72	76	79	10	<0.001
	Ages 51-64	502	81	7	74	79	82	85	88	6	0.012
	Total (Ages 5-64)	1,411	77	7	69	74	78	81	83	7	<0.001
Medicaid	Ages 5-11	921	69	9	59	63	70	75	81	12	<0.001
	Ages 12-18	635	58	10	47	52	58	64	68	12	<0.001
	Ages 19-50	499	47	10	34	41	48	54	58	13	<0.001
	Ages 51-64	238	49	11	35	43	50	56	62	13	<0.001
	Total (Ages 5-64)	1,977	59	9	48	54	61	65	70	11	<0.001

**Includes data submitted by 386 Commercial plans and 164 Medicaid plans to HEDIS for this measure for measurement year 2014*
EP: Eligible Population, the average denominator size across all plans submitting 2014 HEDIS data
IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.e., *what do the results mean in terms of statistical and meaningful differences?*)
2015 Update: The results above indicate there is a 6-13% gap in performance between the 25th and 75th percentile performing plans across the different age ranges and product lines. For all product lines and age groups the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance

is for adults ages 19-50 and 51-64 in Medicaid plans, which show a 13 percentage point gap between 25th and 75th percentile plans.

To put these meaningful differences in performance into context, we estimated that on average 217 additional members per Medicaid plan would meet the asthma medication ratio of 0.50 or higher if plans in the 25th percentile performed as well as plans in the 75th percentile. This estimate is based on the average health plan eligible population.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.e., *what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.e., *what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

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 all 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 followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will 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

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Ratings http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Regulatory and Accreditation Programs NCQA Health Plan Accreditation http://www.ncqa.org/tabid/123/Default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple

	organizations) Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.

HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. [REPLACE or ADD as appropriate, 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives.] Health plans are scored based on performance compared to benchmarks.

QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

From 2012-2014, the measure showed slight improvement (approximately 2 percentage points) across Medicaid health plans (see section 1b.2 for summary of data from health plans). There was also improvement in performance for Medicaid plans at the 90th percentile (+3 percentage points). These data are nationally representative.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of**

initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
More Medicaid plans reported the measure in 2014 compared to 2013 and 2012, which may help explain why the performance rates did not substantially improve. There is hope that with increasing attention to this measure in NCQA's health plan accreditation program, performance rates will begin to improve particularly for Commercial plans.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no identified unintended consequences for this measure during testing or since implementation.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0047 : Asthma: Pharmacologic Therapy for Persistent Asthma
0548 : Suboptimal Asthma Control (SAC) and Absence of Controller Therapy (ACT)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
0047 assesses whether a patient was prescribed controller medication at least once during the measurement year, while 1800 assesses the ratio of controller medications to controller plus reliever medications. There is no impact on interpretability or added burden of data collection because the focus of each measure is different. Also, both measures use value sets to identify asthma controller medications that do not conflict. 0548 is a health plan-level measure that assesses overutilization of rescue medication and need for additional therapeutic intervention. However, 0548 assesses it over a shorter time period (a 90-day period) compared to 1800 (over a measurement year). Also, 1800 assesses a ratio of controller to reliever medications in order to take into account the patients who have severe asthma and may need higher amounts of reliever medication, but still have their asthma under control due to taking daily controller medications.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 1785-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Respiratory Measurement Advisory Panel (RMAP) Members:
David Au, MD, MS, (CHAIR) Associate Prof. of Medicine/Investigator HSRD, Department of Veterans Affairs
Kurt Elward, MD, MPH, Clinical Professor of Family Medicine, Virginia Commonwealth University
Laura Feemster, MD, MS, Assistant Professor of Medicine, University of Washington Medical Center, VA HSR&D
Anne Fuhlbrigge, MD, Clinical Director, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School
Min Joo, MD, MPH, FCCP, Assistant Professor of Medicine, University of Illinois at Chicago/ Jesse Brown VA Medical Center
Christine Joseph, PhD, MPH, BSc, Associate Director of Research, Epidemiologist, Henry Ford Health System
Jerry Krishnan, MD, PhD, Prof. of Medicine & Public Health, Director of Population Health Sciences, AVP, Office of the VP for Health Affairs, University of Illinois Hospital & Health Sciences System
Todd Lee, PharmD, PhD, Primary: Senior Investigator, Secondary: Associate Professor, University of Illinois at Chicago
Allan Luskin, MD, Physician Pulmonologist, Healthy Airways
Richard O´Connor, MD, Director, Dept. of Quality Management, Allergist/Immunologist, Sharp Rees-Stealy Medical Group

Committee on Performance Measurement (CPM) Members:
Peter Bach, MD, Director, Health Policy Center, Memorial Sloan Kettering Cancer Center
Bruce Bagley, MD, FAAFP, Senior Advisor to the Professional Satisfaction and Practice Sustainability effort at the American Medical Association
Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna
Patrick Conway, MD, MSC, Chief Medical Officer and Deputy Administrator, Centers for Medicare and Medicaid Services
Jonathan D. Darer, MD, MPH, Chief Innovation Officer, Geisinger Health System
Helen Darling, Strategic Advisor, National Business Group of Health
Rebekah Gee, MD, MPH, FACOG, Assistant Professor, LSUHSC
Foster Gesten, MD, FACP, New York State Department of Health
Marge Ginsburg, Executive Director, Center for Healthcare Decisions
David Grossman, MD, MPH, Medical Director, Population Health
Christine S. Hunter, MD (Co- Chair), Chief Medical Officer, US Office of Personnel Management
Jeffery Kelman, MMSc, MD, Chief Medical Officer, United Stated Department of Health and Human Services
Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group

J. Brent Pawlecki, MD, MMM, Chief Health Officer, The Goodyear Tire & Rubber Company
Susan Reinhard, PhD, RN, Senior Vice President, AARP Public Policy Institute
Eric C. Schneider, MD, MSc, FACP (Co-chair), Senior Vice President, Policy and Research, The Commonwealth Fund
Marcus Thygeson, MD, MPH, Chief Health Officer, Blue Shield of California

HEDIS EXPERT CODING PANEL Members:
Glen Braden, MBA, CHCA, Attest Health Care Advisors, LLC
Denene Harper, RHIA, American Hospital Association
DeHandro Hayden, BS, American Medical Association
Patience Hoag, RHIT, CPHQ, CHCA, CCS, CCS-P, Health Services Advisory Group
Nelly Leon-Chisen, RHIA, American Hospital Association
Tammy Marshall, LVN, Aetna
Alec McLure, RHIA, CCS-P, Verisk Health
Michele Mouradian, RN, BSN, McKesson Health Solutions
Craig Thacker, RN, CIGNA HealthCare
Mary Jane F. Toomey, RN CPC, Aetna Better Health

Technical Measurement Advisory Panel Members:
Andy Amster, MSPH, Kaiser Permanente
Kathryn Coltin, MPH, Independent Consultant
Lekisha Daniel-Robinson, Centers for Medicare and Medicaid Services
Marissa Finn, MBA, Cigna HealthCare
Scott Fox, MS, MEd, Independence Blue Cross
Carlos Hernandez, CenCalHealth
Kelly Isom, MA, RN, Aetna
Harmon Jordan, ScD, RTI International
Ernest Moy, MD, MPH, Agency for Healthcare Research and Quality
Patrick Roohan, New York State Department of Health
Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC
Natan Szapiro, Independence Blue Cross

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2011
Ad.3 **Month and Year of most recent revision:** 07, 2015
Ad.4 **What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.
Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:** These performance Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.
THE MEASURES AND SEPCIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1893
De.2. **Measure Title:** Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** The measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR), defined as death from any cause within 30 days after the index admission date, for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD. CMS annually reports the measure for patients who are aged 65 or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals.
1b.1. **Developer Rationale:** The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for COPD. 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 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.

Mortality following admission for COPD exacerbation is a priority area for outcomes measure development as it is an outcome that can be influenced by care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices, as well as increase transparency for consumers.

S.4. **Numerator Statement:** The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days from the date of admission for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD.
S.7. **Denominator Statement:** This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 or older or (2) patients aged 40 years or older.

The cohort includes admissions for patients discharged from the hospital with either a principal discharge diagnosis of COPD (see codes below) OR a principal discharge diagnosis of respiratory failure (see codes below) with a secondary discharge diagnosis of acute exacerbation of COPD (see codes below); and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients aged 65 or older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.
S.10. **Denominator Exclusions:** The mortality measures exclude index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
2. Enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission; or
3. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Jan 08, 2013 **Most Recent Endorsement Date:** Jan 08, 2013

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure is paired with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization.

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

The developer reports the following:

- This measure calculates hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization.
- COPD affects as many as 24 million individuals in the United States and is the nation's third leading cause of death.As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care (supplemental oxygen and noninvasive ventilation), discharge planning, management of care transitions, medication reconciliation, and patient education.
- The developer reports numerous studies have demonstrated that appropriate, guideline recommended care and timely treatment for COPD patients can reduce the risk of mortality within 30 days of hospital admission

Guidance from the Evidence Algorithm : 1→2 (eligible for PASS rating)

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The underlying rationale appears to be the same since the last NQF endorsement review. Does the Committee agree and so there is no need for repeat discussion and vote on Evidence?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for

improvement.

- The developer reports aggregate performance data, as follows:

	7/2011 – 6/2012	7/2012 – 6/2013	7/2013 – 6/2014	7/2014 – 6/2015
# hospitals	4538	4513	4496	4658
#admissions	264,721	277,426	238,732	780,879
Mean rate (SD)	7.7 (0.7)	8.1 (08)	7.4 (0.6)	7.8 (0.9)
Range (min-max)	5.5 – 11.3	5.6 – 12.1	4.5 – 12.0	4.8 -13.8
10th percentile	7.0	7.2	6.7	6.7
90th percentile	8.4	9.0	8.2	8.9

Disparities

- The developer provides the following information (July 2011-June 2014):

	#hospitals	*# admissions*	*Minimum rate*	*Median rate*	*Maximum rate*
Dual eligibles					
High proportion	943	142,676	4.8	7.5	11.3
Low proportion	940	212,811	4.9	7.7	13.8
African Americans					
High proportion	937	221,974	4.8	7.5	11.9
Low proportion	1107	100,349	5.1	7.7	12.1
AHRQ SES score					
High proportion	937	152,334	5.0	7.5	11.3
Low proportion	935	180,834	5.1	7.7	13.8

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *What does the data show about disparities?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The developer reports the following:

- This measure calculates hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization.
- COPD affects as many as 24 million individuals in the United States and is the nation's third leading cause of death.As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care (supplemental oxygen and noninvasive ventilation), discharge planning, management of care transitions, medication reconciliation, and patient education.
- The developer reports numerous studies have demonstrated that appropriate, guideline recommended care and timely treatment for COPD patients can reduce the risk of mortality within 30 days of hospital admission

**The evidence provided indicated to a direct outcomes measure. Interventions could contribute to better outcomes.

**Yes, appropriate evidence based care. Appropriate use of NPPV, for example.

Basically agree no change and thus no need to re-vote. Only concern would be ability of risk-adjustment to truly allow for comparison across hospitals.

**The evidence appears to provide good rationale to believe that variation in quality of care will impact the measured thirty day mortality rate. These variations may include differences in therapy, compliance with guidelines, nursing care, airway management, discharge decisions, and post discharge compliance and care

**As in measure 0468:

As a rationale for measuring this health outcome, the developer states hospitals are able to influence mortality rates through a

broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education. I agree that the developer has sufficient evidence to support the measure, however the implementation challenge lies in providers having the processes in place and ability to track those activities to determine where improvement is needed at their respective organizations and the extent of impact when changes are made. How does one determine what is the degree of influence/impact that the aforementioned activities have on mortality rates (e.g., discharge planning reduced by 5% whereas med rec 10% etc., so aiming for changes in activities with the potential for greatest impact on outcomes)

1b. Performance Gap

Comments:

**Yes. The difference between the high and low proportions were significant particularily in the African American and Dual Eligible populations.

**The performance gap was not as clearly demonstrated. Though data did indicate that dual eligible individuals were likely to have worse outcomes.

**Rates look pretty similar across groups. Not sure that there is disparities. Also, wouldn't the differences across race and SES be risk-adjusted?

**There is considerable difference between the highest and lowest measured scores supporting gaps. However, in the reported subgroups which included race=AA, dual eligible, and AHRQ SES score there was little detectable difference so no real evidence of disparities. (example high AA vs low AA median rate = 7.5 vs 7.7)

**Data are provided and calculated from time periods prior to measure implementation to post implementation. Little difference noted in reported rates from pre- and post-implementation however the timing of

1c. High Priority (previously referred to as High Impact)

Comments:

**NA

**n/a

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

- The measure is specified as a facility-level measure for the hospital/acute care setting; the data source for the measure is administrative claims.
- The definition of mortality is "death from any cause within 30 days from the date of admission for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD."
- This measure can be used in either of two patient cohorts: (1) patients aged 65 or older or (2) patients aged 40 years or older.
- Specifications include ICD-9 and ICD-10 codes to identify patients with a COPD or respiratory failure discharge, date of birth, and transfer status, based on admission and discharge dates (used in denominator) and discharge disposition, (used for exclusions). No specific code for identifying death is provided, as the developer has used different datasets (Medicare Enrollment Database; California vital statistics file) in testing of this measure. The developer also notes that other data sources (e.g., the CDC's National Death Index or the SSA's Death Master File could be used to identify date of death).
- ICD-10 codes are included in the specifications. The ICD-9, and ICD-10 codes are described in the numerator and denominator details, and included in the data dictionary attachment. The ICD 9→10 conversion goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

- Only those patients enrolled in both Part A and Part B Medicare for 12 months prior to the index admission are included when the measure is used for the Medicare population. It is unclear, although assumed, that there is a similar 12-month enrollment requirement for an all-payer population.
- If a patient has more than one index admission, only one is randomly chosen for use in the measure.
- Patients are excluded from the measure if age is > 115, the discharge date is prior to the admission date, sex is neither male nor female, admitted to hospice in the 12 months prior to the index admission or on the first day of the index admission, or if discharged against medical advice.
- The calculation algorithm, included in S.18, describes how the risk-standardized mortality ratio is calculated.
- This outcome measure is risk-adjusted using a statistical risk model with 42 factors (age and various co-morbidity indicators).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Testing was done at the measure score level. The previous Committee felt the measure specifications were clear and consistent and can be reliably measured.

Describe any updates to testing: Reliability testing at the measure score level was conducted for a more recent time period (2011-2014)

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- Data used for reliability testing included July 2011-June 2014 Medicare fee-for-service inpatient and outpatient claims; this dataset included information for 780,879 admissions and 4,658 hospitals.
- Developers used a split-sample (or "test-retest") methodology to test score-level reliability. This is an appropriate method. For this analysis, developers randomly assigned half of the patients in each hospital to two separate groups, calculated the performance measure score for each hospital in each of the two groups, and compared the agreement between each hospital's paired scores using the intra-class-correlation coefficient (ICC) and applying a correction factor to account for the overall sample size. The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospitals.
- Although the developer reported assessing data element reliability by comparing model variable frequencies and odds ratios from logistic regression models across the most recent three years of data, NQF does not consider temporal consistency to be a valid method of demonstrating reliability of data elements.

Results of reliability testing

- The ICC values from the split-sample analysis 0.51, indicating that 51% of the variance in scores are due to differences between hospitals. According to the Landis and Koch classification, an ICC value of 51% can be interpreted as moderate agreement. However, a value of 0.7 is often regarded as a minimum acceptable reliability value. Note that the developer states that reliability testing was "was limited to hospitals with 12 or more cases in each split sample", although the measure itself is not limited to hospitals with 12 or more cases..

Guidance from the Reliability Algorithm : 1→2→4→5→6 (eligible for HIGH rating)

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *The reliability testing provided by the developer has not been updated. Does the Committee agree there is no need for repeat discussion and vote on Reliability?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- The developers conducted a systematic assessment of measure face validity by a Technical Expert Panel (TEP) of national experts and stakeholder organizations.

Describe any updates to validity testing: None

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**
☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer assessed the face validity of the measure score as an indicator of quality by soliciting the 11 member TEP's agreement with the following statement: "The risk-standardized mortality rates obtained from the COPD mortality measure as specified will provide an accurate reflection of quality."

Validity testing results:

- Of the TEP members who responded (10 out of 11), 90% agreed (somewhat, moderately, or strongly) that the measure will provide an accurate reflection of quality (60%percent moderately or strongly agreed).

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree the score from this measure as specified is an indicator of quality?*
- *The validity testing provided by the developer has not been updated. Does the Committee agree there is no need for repeat discussion and vote on Validity?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- To ascertain impact of exclusions on the cohort, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion):
  - Age is > 115, discharge date prior to admission date, sex neither male nor female - - <0.01%
  - Enrollment in Medicare hospice - 1.37%
    - During the previous review of this measure CSAC members questioned whether the exclusion is broad enough since the condition of some COPD patients may not be well established in the first 24 hours in order to determine if a hospice or palliative care approach is preferred. The CSAC requested further analyses from the developer and review by the Steering Committee before a final recommendation is made.
    - The Committee agreed that the CSAC raised important questions about palliative care. Committee members noted that the lack of having the end-of-life discussion earlier in the disease process contributes to the problem. The Committee struggled with trying to understand whether there would be a systematic bias between institutions for the timing of hospice referrals. The Committee asked CMS/Yale if they could provide a sensitivity analysis on the effect of extending the hospice exclusions to 2, 3 or 4 days until the end of care.
    - After review of the sensitivity analysis, the majority of Committee members supported its previous evaluation of the scientific acceptability of the measure and maintained its recommendation to endorse the measure with the exclusion of hospice before and on first day of hospitalization.
  - Discharged AMA - 0.58%

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

Description of the model

- The measure is risk-adjusted using a hierarchical logistic regression model with 42 factors to create a hospital-level 30-day risk-standardized mortality rate that simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals using Medicare claims data.
- The model adjusts for case differences based on age and the clinical status of the patient at the time of admission using condition categories (CCs). Only comorbidities that conveyed information about the patient at time of admission or in the 12-months prior, and not complications that arose during the course of the hospitalization, were included in the risk-adjustment.
  - Several datasets have been used in the development of this measure. Initial development of the measure, including the risk-adjustment approach, used development and validation datasets derived from 2008 Medicare claims. The most recent update to the risk-adjustment approach uses Medicare claims from July 1, 2011 – June 30, 2014.
  - The table in section 2b4.4a lists the variables (and associated odds ratios) included the most recent version of the risk-adjustment model.

Performance of the model

- The *c*-statistics for the initial development and validation samples were 0.720 and 0.723, respectively. For the most recent data, the c-statistic was 0.72. The c-statistic is model discrimination statistic that represents the proportion of all-possible pairs with different observed outcomes for which the model correctly predicts a higher

probability for observations with the outcome of interest than those without the outcome of interest. A *c*-statistic of 0.72 means that for 72% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable. .

- Developers also noted the predictive ability of the risk-adjustment model as another indicator of its discrimination. For the most recent dataset, the lowest decile was 1.4% and the highest decile was 21.4%. The spread for the initial development and validation samples was similar, although slightly wider. A wide range between the lowest decile and highest decile indicates an ability to distinguish between high- and low-risk patients.
- The risk-decile plot based on the most recent dataset indicates good model fit (or calibration), as observed values are relatively similar to predicted values across the risk-deciles. Additional calibration statistics from the initial development and validation samples also were provided [(γ0=-0.034, γ1=0.985) and (γ0=0.009, γ1=1.004), respectively]; these also indicate good model fit because the values of γ0 and γ1 are close to 0 and 1, respectively.
- A similar risk-adjustment approach was tested using a California all-payer dataset of patients ages 18+. The developers report good model discrimination and calibration from this approach and therefore consider use of the measure for all-payer patients ages 40+ to be appropriate.

Conceptual basis and empirical support for potential inclusion of SDS factors in risk-adjustment approach

- The developer noted the following SDS factors that have been examined in the mortality literature for various conditions (not just COPD): patient-level self-reported or documented race or ethnicity, income, and education level; median household; Agency for Healthcare Research and Quality (AHRQ)-validated SES index score; and the proportion of Medicaid patients served in the hospital. The developer identified several potential conceptual pathways to consider:
  - Relationship of socioeconomic status (SES) factors or race to health at admission.
  - Use of low-quality hospitals.
  - Differential care within a hospital.
  - Influence of socioeconomic status (SES) on mortality risk outside of hospital quality and health status.
- Based on their interpretation of the literature and analysis of the above pathways, the developers identified 3 potential SDS variables for empirical analysis:
  - African American race (as compared to all others)
  - Dual eligible status
  - AHRQ SES index score (based on 5-digit ZIP code data; includes percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room).
- Analyses indicate that the prevalence of these 3 SDS factors vary across measured entities and are associated with the measured outcome. However, when including any of the 3 SDS variables in a multivariable model that includes all of the claims-based clinical variables, the effect size of each of these variables is small, the c-statistic is similar, and hospital-specific results change little. Moreover, the effect of each of the 3 SDS variable was protective, which is the opposite of what was expected based on the literature reviewed..
- Based on the empirical results, the developer decided NOT to include any of the 3 SDS factors analyzed in the final risk-adjustment model.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure? That is, is the risk model adequate to control for difference in case mix across providers?*
- *Are all of the risk-adjustment variables present at the start of care?*
- *Do you agree with the developer's decision, based on empirical analysis results, to not include the available SDS factors in the risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To demonstrate that meaningful differences between providers can be identified with this measure, the developer reports results as they are designated by CMS on the Hospital Compare website (i.e., based on a 95% interval estimate which could be lower or higher, or could include the national observed rate).
- These results indicate that out of 4,658 hospitals in the U.S., 51 performed "better than the U.S. national rate," 3,611 performed "no different from the U.S. national rate," and 89 performed "worse than the U.S. national rate." Another 907 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

Questions for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Are the measure results meaningful to stakeholder audiences?*

2b6. Comparability of data sources/methods:

- Not Applicable

2b7. Missing Data

- An analysis of missing data analysis was not provided on this measure. Typically, however, there is very little missing data from claims measures.

Guidance from the Validity Algorithm : Box 1→2→3→4→5 (eligible for MODERATE rating)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The are no inconsistancies that I can see.
**The measure specifications were consistent with the evidence and reflected the target population values.
**Yes
**Specifications seem consistent with the evidence
**No challenges identified

2a2. Reliability Testing

Comments:

**The developers conducted a systematic assessment of measure face validity by a Technical Expert Panel (TEP) of national experts and stakeholder organizations.
The developer assessed the face validity of the measure score as an indicator of quality by soliciting the 11 member TEP's agreement with the following statement: "The risk-standardized mortality rates obtained from the COPD mortality measure as specified will provide an accurate reflection of quality."
Of the TEP members who responded (10 out of 11), 90% agreed (somewhat, moderately, or strongly) that the measure will provide an accurate reflection of quality (60%percent moderately or strongly agreed).
**The validity testing was adequate in scope and should be able to be generalized for widespread implementation.
**Maybe. 60% moderately or strongly agreed. Poor validity.
Maybe. Could be impacted by nonclinical factors.
If this is old data that has been reviewed, agree that there is no obvious reason to repeat discussion and vote.
**Only face validity was measured.
**Face validity used, no specifications changes noted. Seems that with implementation data available that additional empirical testing could have also been pursued; however understanding that it is not necessary

2b2. Validity Testing

Comments:

**The exclusions are consistant with the evidence and the risk-adjustment stategy appears sufficient to properly correlate the data provided.
An analysis of missing data analysis was not provided on this measure. Typically, however, there is very little missing data from claims measures.
**No concerns.

**2b3. Yes. No.
2b4. The risk model has good fit and discrimination.
Presumably, having a person enrolled for prior 12 months allows for all risk-adjustment variables to be known. Not sure if this has been tested.
Am curious to learn more about the decision not to include race, SES, or dual eligible status. If these are not important factors in outcome, does that say something about the presence of disparities?
2b5. Results show a bell curve distribution. Not sure that such proves that there are meaningful differences in quality.
No idea of consumers feel this data is meaningful.
2b6. NA
2b7. No issue
**Exclusions seem adequate although there was some debate about hospice exclusions which seems to have been resolved.
The risk adjustment was initially based on 42 factors, but was not able to include 3 SDS factors (AHRQ SES status, risk = AA and dual eligible status) as the statistics and direction of the effect were not consistent with it being able to be included. The overall measure performed only modestly in distinguishing quality across facilities. The results indicate that out of 4,658 hospitals in the U.S., 51 performed "better than the U.S. national rate," 3,611 performed "no different from the U.S. national rate," and 89 performed "worse than the U.S. national rate." Another 907 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
**No perceived challenges

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Testing was done at the measure score level. The previous Committee felt the measure specifications were clear and consistent and can be reliably measured. Reliability testing at the measure score level was conducted for a more recent time period (2011-2014).

- The ICC values from the split-sample analysis 0.51, indicating that 51% of the variance in scores are due to differences between hospitals. According to the Landis and Koch classification, an ICC value of 51% can be interpreted as moderate agreement. However, a value of 0.7 is often regarded as a minimum acceptable reliability value. Note that the developer states that reliability testing was "was limited to hospitals with 12 or more cases in each split sample", although the measure itself is not limited to hospitals with 12 or more cases.

**Reliability testing was completed with an adequate sample size. The testing was conducted at the measure score level.
**2. Poor
Yes, test sample was adequate.
No, ICC of 0.51 seems too low.
If this is old data that has been reviewed, agree that there is no obvious reason to repeat discussion and vote.
**Reliability testing was conducted at the measure score level, and has been done previously and also repeated with a more recent cohort using a split sample test / retest approach. the results of this test wee only moderate (ICC value of 0.51; 0.7 being the acceptable minimum. The reliability of the measure cannot be considered high.
**The random assignment of an index admission when there are multiple relevant admissions poses challenges in consistency of implementation, however overall reliability testing at score level were sufficient without considering this.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.
As such, I have no concerns about this strategy being put into operational use.
**The data elements are administrative and E.H.R. fields. No concerns with the feasibility of the measure.
**Poor -- no mention of death assessment
**Feasibility should be good; all data elements available electronically.
**Concerns regarding death date and availability only in sources external to claims data.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
- This measure is publically reported nationally on Hospital Compare.

Improvement results :
- The developer reports aggregate performance data, as follows:

	7/2011 – 6/2012	7/2012 – 6/2013	7/2013 – 6/2014	7/2014 – 6/2015
# hospitals	4538	4513	4496	4658
#admissions	264,721	277,426	238,732	780,879
Mean rate (SD)	7.7 (0.7)	8.1 (08)	7.4 (0.6)	7.8 (0.9)
Range (min-max)	5.5 – 11.3	5.6 – 12.1	4.5 – 12.0	4.8 -13.8
10th percentile	7.0	7.2	6.7	6.7
90th percentile	8.4	9.0	8.2	8.9

The developer states there has been significant progress in 30-day RSMR for COPD:
- The median 30-day RSMR decreased by 0.2 absolute percentage points from July 2011-June 2012 (median RSMR: 7.6%) to July 2013-June 2014 (median RSMR: 7.4%).
- The median hospital RSMR from July 2011-June 2014 was 7.7% (IQR 7.2% - 8.2%).
- In addition, hospitals with a high proportion of dual eligible and African American patients achieve a similar range of performance as compared with hospitals with a low proportion of these patients. Also, hospitals with a low proportion of patients with AHRQ SES index scores equal to or below 45.1 perform similarly to hospitals with a high proportion of patients with AHRQ SES index scores equal to or below 45.1. These results indicate that both groups of hospitals can perform well on the measure.

Unexpected findings (positive or negative) during implementation: Developer states there were no unexpected findings during implementation.

Potential harms: There were no identified unintended consequences for this measure during testing or since implementation.

Feedback : During the 2014-2015 MAP review, the MAP supported this measure for inclusion in the VBP program. The group agreed that the program currently includes 30 day mortality rates for AMI, heart failure, and pneumonia, thus including 30 day mortality rates for COPD would be appropriate.

***Questions for the Committee**:*

- *Has there been improvement in measure results?*
- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**The current uses of the measure are publicly reported
and are being used in an accountability program.
Specifically, this measure is publically reported nationally on Hospital Compare.

**Yes, there has been improvement
Potentially can be used to further goal of high quality care.
I believe the benefits outweigh the risks.
**The measure is currently being publicly reported. The trend in improvement suggests that it might be having an impact on improving the quality of care in measured facilities. There do not seem to be any implementation issues or unintended consequences.
**Measure is useful, however largely impossible for hospitals to attempt to replicate and implement at institutional level to monitor performance and make more real-time advances to improve. Underlying processes shown to impact pneumonia mortality can be implemented and tracked but extent of impact on improving pneumonia outcomes would be difficult to discern outside of the infrequent CMS calculation and reporting of claims roll-up.

Criterion 5: Related and Competing Measures

Related or competing measures

- 1891 : Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization

Harmonization

- Developer response: We did not include in our list of related measures any non-outcome (for example, process) measures with the same target population as our measure. Our measure cohort was heavily vetted by clinical experts, a technical expert panel, and a public comment period. Additionally, the measure, with the specified cohort, has been publicly reported since December 2014. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1893

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

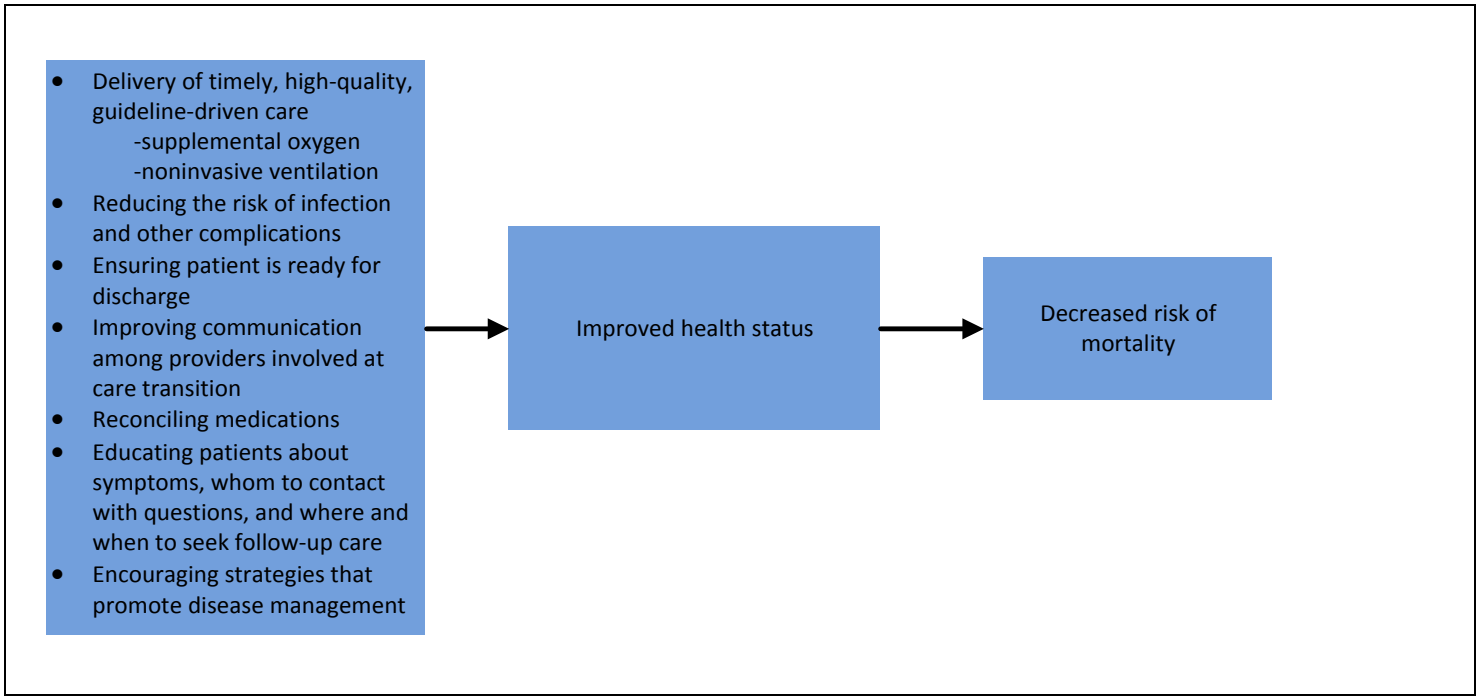
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized mortality rates following hospitalization for COPD. 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 mortality 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.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

COPD is a priority condition for outcomes measure development because it is a leading cause of morbidity and mortality. COPD affects as many as 24 million individuals in the United States and is the nation's third leading cause of death (National Heart, Lung, and Blood Institute, 2009; CDC, 2011; Wier et al., 2011; CDC, 2013; American Lung Association, 2015). Studies report that in-hospital mortality rates for patients hospitalized for exacerbations of COPD range from 2-5% (AHRQ, National Statistics on All Stays; Patil et al., 2009; Tabak et al., 2009; Lindenauer et al., 2006; Dransfield et al., 2008) and 30-day mortality rates range from 3-9% (Faustini et al., 2008; Fruchter et al., 2008; Lindenauer et al., 2013).

In 2011 COPD was one of the top 20 most expensive conditions treated in U.S. hospitals (AHRQ, 2011). It was also one of the top 20 most expensive conditions billed to Medicare, accounting for nearly $4,074,000 of total hospital charges billed to Medicare (AHRQ, 2011).

Many current hospital processes have been associated with lower mortality rates within 30 days of hospital admission (Jha et al., 2007). In COPD in particular, supplemental oxygen and the use of noninvasive ventilation in carefully selected patients has been shown to improve both short and long term survival. Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. Measurement of patient outcomes, such as mortality, allows for a comprehensive view of quality of care that reflects complex aspects of care such as communication between providers and coordinated transitions to the outpatient environment. These aspects are critical to patient outcomes, and are broader than what can be captured by individual process-of-care measures.

The COPD RSMR measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Krumholz et al., 2007).

The diagram above indicates some of the many care processes that can influence mortality risk. Numerous studies have demonstrated that appropriate (guideline recommended care) and timely treatment for COPD patients can reduce the risk of mortality within 30 days of hospital admission (Krumholz et al., 2007; Williams et al., 2012; Ram et al., 2004; Austin et al., 2010; Ntoumenopoulos, 2011).

References:

Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf. Accessed November 16, 2015.

Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). National Statistics on All Stays. Available at: http://hcupnet.ahrq.gov./HCUPnet.jsp?Id=DDB2B43489A1ADB8&Form=DispTab&JS=Y&Action=Accept. Accessed December 4, 2015.

Austin MA, Wills KE, Blizzard L, et al. Effect of high flow oxygen on mortality in chronic obstructive pulmonary disease patients in prehospital setting: randomised controlled trial. BMJ. 2010 Oct 18; 341:c5462. doi: 10.1136/bmj.c5462.

American Lung Association. Lung Health & Diseases. Learn About Chronic Obstructive Pulmonary Disease (COPD)—How Serious is COPD, 2015. Available at: http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/copd/learn-about-copd/how-serious-is-copd.html. Accessed November 16, 2015.

Centers for Disease Control and Prevention. Detailed Tables for the National Vital Statistics Report (NVSR) "Deaths: Final Data for 2013". http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf Accessed November 16, 2015.

Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report (MMWR). Chronic Obstructive Pulmonary Disease Among Adults — United States, 2011. November 23, 2012/ 61(46); 938-943. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6146a2.htm?s_cid=mm6146a2_w. Accessed November 16, 2015.

Dransfield MT, Rowe SM, Johnson JE, et al. Use of beta blockers and the risk of death in hospitalized patients with acute exacerbations of COPD. Thorax. Apr 2008; 63 (4): 301-305.

Faustini A, Marino C, D'Ippoliti D, et al. The Impact on risk-factor analysis of different mortality outcomes in COPD patients. European Respiratory Journal. Sep 2008; 32 (3): 629-636.

Fruchter O, Yigla M., Predictors of long-term survival in elderly patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Respirology. Nov 2008; 13 (6): 851-855.

Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. Health Aff (Millwood) 2007 Jul-Aug; 26(4):1104-10.

Krumholz HM, Normand SL, Spertus JA, et al. Measuring performance for treating heart attacks and heart failure: the case for outcomes measurement. Health Aff (Millwood). 2007 Jan-Feb;26(1):75-85.

Lindenauer PK, Grosso LM, Wang C, et al. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. J Hosp Med. 2013 Aug; 8(8):428-35. doi: 10.1002/jhm.2066. Epub 2013 Jul 26.

Lindenauer PK, Pekow P, Gao S, et al. Quality of care for patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Ann Intern Med. June 20 2006; 144 (12): 894-903.

National Heart, Lung, and Blood Institute. Morbidity & Mortality: Chart Book on Cardiovascular, Lung, and Blood Diseases. 2009. Available at: http://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook_508.pdf. Accessed November 16, 2015.

Ntoumenopoulos G. Using titrated oxygen instead of high flow oxygen during an acute exacerbation of chronic obstructive pulmonary disease (COPD) saves lives. J Physiother. 2011; 57(1):55. doi: 10.1016/S1836-9553(11)70008-X.

Patil SP, Krishnan JA, Lechtzin n, Diette GB. In-hospital mortality following acute exacerbations of chronic obstructive pulmonary disease. Archives of Internal Medicine. Sep 28 2009; 169 (17): 1595-1602.

Ram FS, Picot J, Lightowler, Wedzicha JA. Non-invasive positive pressure ventilation for treatment of respiratory failure due to exacerbations of chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2004; (1):CD004104.

Tabak YP, Sun X, Johannes RS, et al. Mortality and need for mechanical ventilation in acute exacerbations of chronic obstructive pulmonary disease: development and validation of simple risk score. Archives of Internal Medicine. Sep 28 2009; 169(17): 1595-1602.

Wier LM, Elixhauser A, Pfuntner, et al. Overview of Hospitalizations among Patients with COPD, 2008. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs #106 [Internet]. Rockville (MD): Agency for Health Care Policy and Research (US). Feb. 2011. Available at: http://www.ncbi.nlm.nih.gov/books/NBK53969/. Accessed November 16, 2015.

Williams JW, Cox CE, Hargett CW, et al. Noninvasive Positive-Pressure Ventilation (NPPV) for Acute Respiratory Failure [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Jul. Report No.: 12-EHC089-EF. AHRQ Comparative Effectiveness Reviews.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
COPD_Mortality_Measure_Evidence_Form_v1.0.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for COPD. 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 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.

Mortality following admission for COPD exacerbation is a priority area for outcomes measure development as it is an outcome that can be influenced by care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices, as well as increase transparency for consumers.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Distribution of Hospital COPD RSMRs over Different Time Periods
Results for each data year

Characteristic	07/2011-06/2012	07/2012-06/2013	07/2013-06/2014	07-2011-06/2014
Number of Hospitals	4,538	4,513	4,496	4,658
Number of Admissions	264,721	277,426	238,732	780,879
Mean (SD)	7.7 (0.7)	8.1 (0.8)	7.4 (0.6)	7.8 (0.9)
Range (min. – max.)	5.5-11.3	5.6-12.1	4.5-12.0	4.8-13.8
Minimum	5.5	5.6	4.5	4.8
10th percentile	7.0	7.2	6.7	6.7
20th percentile	7.2	7.5	7.0	7.1
30th percentile	7.4	7.7	7.2	7.3
40th percentile	7.5	7.9	7.3	7.5
50th percentile	7.6	8.0	7.4	7.7
60th percentile	7.7	8.1	7.5	7.8
70th percentile	7.9	8.3	7.6	8.1
80th percentile	8.1	8.6	7.8	8.4
90th percentile	8.4	9.0	8.2	8.9
Maximum	11.3	12.1	12.0	13.8

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Distribution of COPD RSMRs by Proportion of Dual Eligible Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic	Hospitals with a low proportion (=12.9%) Dual Eligible patients	Hospitals with a high proportion (=26.8%) Dual Eligible patients
Number of Measured Hospitals	943	940
Number of Patients	212,811 patients in low-proportion hospitals	142,676 in high-proportion hospitals
Maximum	13.8	11.3
90th percentile	9.1	8.8
75th percentile	8.3	8.1
Median (50th percentile)	7.7	7.5
25th percentile	7.1	7.0
10th percentile	6.6	6.5
Minimum	4.9	4.8

Distribution of COPD RSMRs by Proportion of African-American Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low proportion (=0.0%) African-American patients//Hospitals with a high proportion (=9.1%) African-American patients
Number of Measured Hospitals// 1,107 // 937
Number of Patients// 100,349 patients in low-proportion hospitals// 221,974 in high-proportion hospitals
Maximum// 12.1 // 11.9
90th percentile// 9.0 // 8.9
75th percentile// 8.3 // 8.2
Median (50th percentile)// 7.7 // 7.5
25th percentile// 7.2 // 6.9
10th percentile// 6.9 // 6.4
Minimum // 5.1 // 4.8

Distribution of COPD RSMRs by Proportion of patients with AHRQ SES Index Scores Equal to or Below 45.1:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims and the American Community Survey (2008-2012) data

Characteristic// Hospitals with a low proportion of patients with AHRQ SES index score equal to or below 45.1 (=2.9%)//Hospitals with a high proportion of patients with AHRQ SES index score equal to or below 45.1 (=44.7%)
Number of Measures Hospitals// 935 // 937
Number of Patients// 180,834 patients in hospitals with low proportion of patients with AHRQ SES index score equal to or below 45.1 // 152,334 patients in hospitals with high proportion of patients with AHRQ SES index score equal to or below 45.1
Maximum// 13.8 // 11.3
90th percentile// 9.1 // 8.8
75th percentile// 8.4 // 8.1
Median (50th percentile)// 7.7 // 7.5
25th percentile// 7.2 // 7.0
10th percentile// 6.7 // 6.6
Minimum // 5.1 // 5.0

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
COPD is a priority condition for outcomes measure development because it is a leading cause of morbidity and mortality. COPD affects at least 15 and as many as 24 million individuals in the United States, and is the nation's third leading cause of death (National Heart, Lung, and Blood Institute, 2009; CDC, 2011; Wier et al., 2011; CDC, 2013; American Lung Association, 2015). Studies report

that in-hospital mortality rates for patients hospitalized for exacerbations of COPD range from 2-5% (AHRQ, National Statistics on All Stays; Patil et al., 2009; Tabak et al., 2009; Lindenauer et al., 2006; Dransfield et al., 2008) and 30-day mortality rates range from 3-9% (Faustini et al., 2008; Fruchter et al., 2008; Lindenauer et al., 2013).

In 2011 COPD was one of the top 20 most expensive conditions treated in U.S. hospitals (AHRQ, 2011). It was also one of the top 20 most expensive conditions billed to Medicare, accounting for nearly $4,074,000 of total hospital charges billed to Medicare (AHRQ, 2011).

1c.4. Citations for data demonstrating high priority provided in 1a.3

Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf. Accessed November 16, 2015.

Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). National Statistics on All Stays. Available at: http://hcupnet.ahrq.gov./HCUPnet.jsp?Id=DDB2B43489A1ADB8&Form=DispTab&JS=Y&Action=Accept. Accessed December 4, 2015.

American Lung Association. Lung Health & Diseases. Learn About Chronic Obstructive Pulmonary Disease (COPD)—How Serious is COPD, 2015. Available at: http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/copd/learn-about-copd/how-serious-is-copd.html. Accessed November 16, 2015.

Centers for Disease Control and Prevention. Detailed Tables for the National Vital Statistics Report (NVSR) "Deaths: Final Data for 2013". Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf. Accessed November 16, 2015.

Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report (MMWR). Chronic Obstructive Pulmonary Disease Among Adults — United States, 2011. November 23, 2012/ 61(46); 938-943. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6146a2.htm?s_cid=mm6146a2_w. Accessed November 16, 2015.

Dransfield MT, Rowe SM, Johnson JE, et al. Use of beta blockers and the risk of death in hospitalized patients with acute exacerbations of COPD. Thorax. Apr 2008; 63 (4): 301-305.

Faustini A, Marino C, D'Ippoliti D, et al. The Impact on risk-factor analysis of different mortality outcomes in COPD patients. European Respiratory Journal. Sep 2008; 32 (3): 629-636.

Fruchter O, Yigla M., Predictors of long-term survival in elderly patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Respirology. Nov 2008; 13 (6): 851-855.

National Heart, Lung, and Blood Institute. Morbidity & Mortality: Chart Book on Cardiovascular, Lung, and Blood Diseases. 2009. Available at: http://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook_508.pdf. Accessed November 16, 2015.

Lindenauer PK, Grosso LM, Wang C, et al. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. J Hosp Med. 2013 Aug; 8(8):428-35. doi: 10.1002/jhm.2066. Epub 2013 Jul 26.

Lindenauer PK, Pekow P, Gao S, et al. Quality of care for patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Ann Intern Med. June 20 2006; 144 (12): 894-903.

Patil SP, Krishnan JA, Lechtzin n, Diette GB. In-hospital mortality following acute exacerbations of chronic obstructive pulmonary disease. Archives of Internal Medicine. Sep 28 2009; 169 (17): 1595-1602.

Tabak YP, Sun X, Johannes RS, et al. Mortality and need for mechanical ventilation in acute exacerbations of chronic obstructive pulmonary disease: development and validation of simple risk score. Archives of Internal Medicine. Sep 28 2009; 169(17): 1595-1602.

Wier LM, Elixhauser A, Pfuntner, et al. Overview of Hospitalizations among Patients with COPD, 2008. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs #106 [Internet]. Rockville (MD): Agency for Health Care Policy and Research (US). Feb. 2011. Available at: http://www.ncbi.nlm.nih.gov/books/NBK53969/. Accessed November 16, 2015.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A. This measure is not a PRO-PM.

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD), Pulmonary/Critical Care : Dyspnea

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Safety, Safety : Complications, Safety : Healthcare Associated Infections

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_1893_S2b_Mortality_Data_Dictionary_v0.3_forCMS.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Annual Updates
1. Updated CC map.

a. Rationale: The ICD-9-CM CC map is updated annually to capture all relevant comorbidities coded in patient administrative claims data.

No other updates or changes have been made since the last endorsement except for use of new years of data for public reporting.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days from the date of admission for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*

Numerator time window: We define the time period for death from any cause within 30 days from the date of admission for the index COPD hospitalization.

Denominator time window: This measure was developed with 12 months of data. The time window can be specified from one to three years. Currently, the measure is publicly-reported with three years of index admissions.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
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 are using this field to define the outcome.

The measure counts deaths for any cause within 30 days of the date of admission of the index COPD hospitalization.

Identifying deaths in the FFS measure
As currently reported, we identify deaths for FFS Medicare patients aged 65 or older in the Medicare Enrollment Database (EDB).

Identifying deaths in the all-payer measure
For the purposes of development of an all-payer measure, deaths were identified using the California vital statistics data file. Nationally, post-discharge deaths can be identified using an external source of vital status, such as the Social Security Administration's Death Master File (DMF) or the Centers for Disease Control and Prevention's National Death Index (NDI).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 or older or (2) patients aged 40 years or older.

The cohort includes admissions for patients discharged from the hospital with either a principal discharge diagnosis of COPD (see codes below) OR a principal discharge diagnosis of respiratory failure (see codes below) with a secondary discharge diagnosis of acute exacerbation of COPD (see codes below); and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients aged 65 or older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:

1. Principal discharge diagnosis of COPD or principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of COPD with exacerbation
2. Enrolled in Medicare fee-for-service (FFS)
3. Aged 65 or over
4. Not transferred from another acute care facility
5. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission, and enrolled in Part A during the index admission.

This measure can also be used for an all-payer population aged 40 years and older. We have explicitly tested the measure in both patients aged 40 years and older and those aged 65 years or older (see Testing Attachment for details).

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:
ICD-9-CM codes used to define COPD:
491.21 Obstructive chronic bronchitis with (acute) exacerbation
491.22 Obstructive chronic bronchitis with acute bronchitis
491.8 Other chronic bronchitis
491.9 Unspecified chronic bronchitis
492.8 Other emphysema
493.20 Chronic obstructive asthma, unspecified
493.21 Chronic obstructive asthma with status asthmaticus
493.22 Chronic obstructive asthma with (acute) exacerbation
496 Chronic airway obstruction, not elsewhere classified
518.81 Acute respiratory failure (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])
518.82 Other pulmonary insufficiency, not elsewhere classified (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])
518.84 Acute and chronic respiratory failure (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])
799.1 Respiratory arrest (Principal diagnosis when combined with a secondary diagnosis of COPD with exacerbation [491.21, 491.22, 493.21, or 493.22])

ICD-9-CM codes used to define acute exacerbation of COPD:
491.21 Obstructive chronic bronchitis with (acute) exacerbation
491.22 Obstructive chronic bronchitis with acute bronchitis
493.21 Chronic obstructive asthma with status asthmaticus
493.22 Chronic obstructive asthma with (acute) exacerbation

ICD-10-CM codes used to define COPD:
J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation
J44.0 Chronic obstructive pulmonary disease with acute lower respiratory infection
J41.8 Mixed simple and mucopurulent chronic bronchitis
J42 Unspecified chronic bronchitis
J43.9 Emphysema, unspecified
J44.9 Chronic obstructive pulmonary disease, unspecified
J96.00 Acute respiratory failure, unspecified whether with hypoxia or hypercapnia
J96.90 Respiratory failure, unspecified, unspecified whether with hypoxia or hypercapnia
J80 Acute respiratory distress syndrome
J96.20 Acute and chronic respiratory failure, unspecified whether with hypoxia or hypercapnia
R09.2 Respiratory arrest

ICD-10-CM codes used to define acute exacerbation of COPD:
J44.1 Chronic obstructive pulmonary disease with (acute) exacerbation
J44.0 Chronic obstructive pulmonary disease with acute low respiratory infection

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
The mortality measures exclude index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
2. Enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission; or
3. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is

randomly selected for inclusion in the cohort.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1. Inconsistent vital status or unreliable data are identified if any of the following conditions are met 1) the patient's age is greater than 115 years: 2) if the discharge date for a hospitalization is before the admission date; 3) if the patient has a sex other than 'male' or 'female'.

2. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice data.

3. Discharges against medical advice (AMA) are identified using the discharge disposition indicator.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
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 measure employs a hierarchical logistic regression model to create a hospital-level 30-day RSMR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of mortality within 30 days of admission for age and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of mortality at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were expected to be predictive of mortality, 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 claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using both inpatient and outpatient Medicare FFS claims data. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and 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 (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). 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 index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care when they are only recorded in the index admission.

The final set of risk adjustment variables is:

Demographics
Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts.

Comorbidities
Sleep apnea (ICD-9 codes 327.20, 327.21, 327.23, 327.27, 327.29, 780.51, 780.53, 780.57)
History of mechanical ventilation (ICD-9 codes 93.90, 96.70, 96.71, 96.72)
Respirator dependence/respiratory failure (CC 77-78)
Cardio-respiratory failure or shock (CC 79)
Congestive heart failure (CC 80)
Coronary atherosclerosis or angina (CC 83-84)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)
Vascular or circulatory disease (CC 104-106)
Fibrosis of lung or other chronic lung disorders (CC 109)
Asthma (CC 110)
Pneumonia (CC 111-113)
Pleural effusion/pneumothorax (CC 114)
Other lung disorders (CC 115)
Metastatic cancer or acute leukemia (CC 7)
Lung, upper digestive tract, and other severe cancers (CC 8)
Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other cancers and tumors; other respiratory and heart neoplasms (CC 9-11)
Other digestive and urinary neoplasms (CC 12)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Other endocrine/metabolic/nutritional disorders (CC 24)
Other gastrointestinal disorders (CC 36)
Osteoarthritis of hip or knee (CC 40)
Other musculoskeletal and connective tissue disorders (CC 43)
Iron deficiency or other unspecified anemias and blood disease (CC 47)
Dementia or other specified brain disorders (CC 49-50)
Drug/alcohol abuse, without dependence (CC 53)
Other psychiatric disorders (CC 60)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Mononeuropathy, other neurological conditions/injuries (CC 76)
Hypertension and hypertensive disease (CC 90-91)
Stroke (CC 95-96)
Retinal disorders, except detachment and vascular retinopathies (CC 121)
Other eye disorders (CC 124)
Other ear, nose, throat and mouth disorders (CC 127)
Renal failure (CC 131)
Decubitus ulcer or chronic skin ulcer (CC 148-149)
Other dermatological disorders (CC 153)
Trauma (CC 154-156, 158-161)
Vertebral fractures (CC 157)
Major complications of medical care and trauma (CC 164)

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.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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 measure estimates hospital-level 30-day all-cause RSMRs following hospitalization for COPD using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of index admission using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution 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 RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on 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 mortality rates or better quality, and a higher ratio indicates higher-than-expected mortality rates or worse quality.

The "predicted" number of deaths (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of mortality. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of deaths (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed mortality rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2011).

Reference:
Grosso L, Lindenauer P, Wang C, et al. Hospital-level 30-day Mortality Following Admission for an Acute Exacerbation of Chronic Obstructive Pulmonary Disease. 2011.

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

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. This measure is not based on a sample.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A. This measure is not based on a survey or patient-reported data.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing values are rare among variables used from claims data in this measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Data sources for the Medicare FFS measure:
1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician 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 data source 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 et al., 1992).

3. The American Community Survey (2008-2012): The American Community Survey data is collected annually and an aggregated 5-years data was used to calculate the AHRQ SES composite index score.

4. Data sources for the all-payer testing: For our analyses to examine use in all-payer data, we used all-payer data from California. 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, we performed analyses to determine whether the COPD mortality measure can be applied to all adult patients, including not only FFS Medicare patients aged 65 or over, but also non-FFS Medicare patients aged 18-64 years at the time of admission.

Reference:
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.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A. This measure is not a composite performance measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
COPD_Mortality_Measure_Testing_Form_v1.1.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 1893
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization
Date of Submission: 12/14/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation

criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☒ other: Census Data/American Community Survey

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The datasets used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). Additionally, census as well as claims data were used to assess socioeconomic factors and race (dual eligible and African American race variables obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score obtained through census data). The dataset used varies by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing?

The dates used vary by testing type; see Section 1.7 for details.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. All non-federal, acute care inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years or over are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of admissions/patients varies by testing type: see Section 1.7 for details

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2)
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65 years or over in the most recent 3-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples. In each year of measure reevaluation, we also re-fit the model and compared the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (**Dataset 1** below).

Dataset 1 (2015 public reporting cohort, version 4.0): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014
Number of Admissions: 780,879
Patient Descriptive Characteristics: average age=77.0, % male=41.5
Number of Measured Hospitals: 4,658

For validity testing (Section 2b2): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
To create the model development and validation samples (**Dataset 2**), we applied the inclusion and exclusion criteria to all 2008 admissions. We randomly selected half of all COPD admissions in 2008 that met the inclusion and exclusions criteria to create a model development sample and used the remaining admissions as our model validation sample.

Dataset 2 (original measure development and validation samples)
Date of Data: 2008

First half of split sample (development sample)
-Number of Admissions: 150,035
-Number of Measured Hospitals: 4,537

Second half of split sample (validation sample)
-Number of Admissions: 149,646
-Number of Measured Hospitals: 4,535

For testing of measure exclusions (Section 2b3)
Dataset 1

For testing of measure risk adjustment (Section 2b4)
Dataset 1

Dataset 2 (development dataset): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims

Dataset 3 (all payer dataset): California Patient Discharge Data

Dates of Data: January 1, 2006 – December 31, 2006
Number of Admissions: 39,232 (all 18+ total); 16,629 (FFS 65+); 9,917 (non-FFS 65+); 12,686 (all 18-64)
Patient Descriptive Characteristics: mean age=69.95 (all 18+ total); mean age=77.49 (FFS 65+); mean age=77.13 (non-FFS 65+); mean age=54.46 (all 18-64)
Number of Measured Hospitals: >450 non-Federal acute care hospitals

The measure was applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations. In addition, the unique patient ID number is used to link with state vital statistics records to assess 30-day mortality.

For testing to identify meaningful differences in performance (Section 2b5)
Dataset 1

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.3)
Dataset 1 and **Dataset 4**: The American Community Survey (2008-2012)

We examined disparities in performance according to the proportion of patients in each hospital who were of African-American race and the proportion who were dual eligible for both Medicare and Medicaid insurances. We also used the AHRQ SES index score to study the association between performance measures and socioeconomic status.

Data Elements

- African-American race and dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from CMS enrollment data (**Dataset 1**)
- Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 4**)

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Sociodemographic status incorporates socioeconomic variables as well as race into a more concise term. However, given the fact that socioeconomic risk factors are distinct from race and should be interpreted differently, we have decided to keep "socioeconomic status" and "race" as separate terms.

We selected socioeconomic status (SES) and race variables to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various SES factors and African-American race to worse health status and higher mortality over a lifetime (Adler and Newman, 2002; Mackenbach et al., 2000; Tonne et al., 2005; van Oeffelen et al., 2012). Income, education, and occupational level are the most commonly examined variables. However, literature directly examining how different SES factors or race might influence the likelihood of older, insured, Medicare patients of dying within 30 days of an admission for COPD is much more limited. The causal pathways for SES and race variable selection are described below in Section 2b4.3.

The SES and race variables used for analysis were:

- Dual eligible status (**Dataset 1**)
- African-American race (**Dataset 1**)
- AHRQ-validated SES index score using 5-digit zip code data (percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room) (**Dataset 4**)

References:

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health affairs (Project Hope).* 2002; 21(2):60-76.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation. Cardiovascular quality and outcomes. May 2014; 7(3):391-397.

Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008; 2.

Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. *European heart journal.* 2000; 21(14):1141-1151.

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. *Circulation.* Jun 14 2005; 111(23):3063-3070.

van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. *European journal of epidemiology.* Aug 2012; 27(8):605-613.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data Element Reliability

In constructing the measure, 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 from logistic regression models across the most recent three years of data (**Dataset 1**).

Measure Score 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 finally comparing the agreement between the two resulting performance measures 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 (ICC) (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used dataset 1 split sample and calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals using the intra-class correlation as defined by ICC (2,1) by Shrout and Fleiss (1979).

Using two independent samples provides a stringent 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. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman, 1910; Brown, 1910). We use this to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

References:
Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

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

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; 86:420-428.

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element reliability results (**Dataset 1**)
The frequency of some model variables increased and others decreased between 2011 and 2014, which may reflect an increase or decrease in the rate of specific comorbidities in the FFS population. For example, there was a notable increase in percent frequency for “other psychiatric disorders” (CC 60) (25.5% to 31.0%), “cardio-respiratory failure or shock (CC 79)” (32.1% to 35.5%), and “sleep apnea” (ICD-9 codes 327.20, 327.21, 327.23, 327.27, 327.29, 780.51, 780.53, and 780.57) (16.7% to 19.0%). There was a notable decrease in percent frequency for “other lung disorders” (CC 115) (52.1% to 48.7%), “fibrosis of lung or other chronic lung disorders” (CC 109) (17.0% to 14.7%), and “coronary atherosclerosis or angina” (CC 83-84) (53.4% to 51.9%). Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.

For the model variable frequencies and risk variable odds ratios, see the *2015 Measure Updates and Specifications Report* (Dorsey et al. 2015) posted on the web page provided in data field S.1.

Measure Score Reliability Results (**Dataset 1**)
There were 780,879 admissions in the combined 3-year sample, with 389,235 in one randomly selected sample and 391,644 in the other sample. The agreement between the two RSMRs for each hospital was 0.51, which according to the conventional interpretation is “moderate” (Landis and Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data.

Reference:
Dorsey K, Grady J, Desai N, et al. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measures Acute Myocardial Infarction –Version 9.0, Heart Failure –Version 9.0, Pneumonia –Version 9.0, Chronic Obstructive Pulmonary Disease –Version 4.0, Stroke –Version 4.0.

https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 5, 2015.

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

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The stability over time of the risk factor frequencies and odds ratios suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates moderate agreement of measure scores across samples using a conservative approach to assessment.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Measure validity is demonstrated through prior validity testing done on our claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP) of national experts and stakeholder organizations.

Validity of Claims-Based Measures

Our team has demonstrated for a number of prior measures the validity of claims-based measures for profiling hospitals by comparing either the measure results or individual data elements against medical records. CMS validated seven NQF-endorsed measures currently in public reporting (AMI, heart failure, and pneumonia mortality and readmission and coronary artery bypass graft surgery or CABG readmission) with models that used chart-abstracted data for risk-adjustment. Specifically, claims model validation was conducted by building comparable models using abstracted medical chart data for risk-adjustment for heart failure patients (National Heart Failure data) (Krumholz et al., 2006 [3]; Keenan et al., 2008), AMI patients (Cooperative Cardiovascular Project data) (Krumholz et al., 2006 [2]), pneumonia patients (National Pneumonia Project dataset) (Bratzler et al., 2011), and CABG patients (Shahian et al., 2014; Suter et al., 2014). When both models were applied to the same patient population, the hospital risk-standardized rates estimated using the claims-based risk adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based models for public reporting.

We have also completed two national, multi-site validation efforts for two procedure-based complications measures (for primary elective hip/knee arthroplasty and implantable cardioverter defibrillator [ICD]). Both projects demonstrated strong agreement between complications coded in claims and abstracted medical chart data.

Validity Indicated by Established Measure Development Guidelines:

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes

measurement set forth in NQF guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006 [1]).

Validity as Assessed by External Groups:
Throughout measure development, we obtained expert and stakeholder input via three mechanisms: regular discussions with an advisory working group, a national Technical Expert Panel (TEP), and a 30-day public comment period in order to increase transparency and to gain broader input into the measure.

The working group was assembled, and regular meetings were held throughout the development phase. The working group was tailored for development of this measure and consisted of three physicians who are board-certified in pulmonary and critical care medicine and a pharmacoepidemiologist with expertise in COPD. All members have expertise in quality measure development. The working group meetings addressed key issues related to measure development, including weighing the pros and cons of and finalizing key decisions (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In addition to the working group, and in alignment with the CMS MMS, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives, including physicians, consumers, and purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held three structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site: https://www.cms.gov/MMS/17_CallforPublicComment.asp. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and contributed to minor modifications to the measure.

Face Validity as Determined by TEP:
One means of confirming the validity of this measure was face validity assessed by our Technical Expert Panel (TEP), which included 11 members including individuals with diverse perspectives and backgrounds, including clinicians, consumers, hospitals, purchasers, and experts in quality improvement.

List of TEP Members
-Darlene Bainbridge, MS, NHA, CPHQ, CPHRM (President/CEO, Darlene D. Bainbridge & Associates, Inc.)
-Robert A. Balk, MD (Director of Pulmonary and Critical Care Medicine, Rush University Medical Center)
-Dale Bratzler, DO, MPH (President and CEO, Oklahoma Foundation for Medical Quality)
-Scott Cerreta, RRT (Director of Education, COPD Foundation)
-Gerard J. Criner, MD (Director of Temple Lung Center and Divisions of Pulmonary and Critical Care Medicine, Temple University)
-Guy D'Andrea, MBA (President, Discern Consulting)
-Jonathan Fine, MD (Director of Pulmonary Fellowship, Research and Medical Education, Norwalk Hospital)
-David Hopkins, MS, PhD (Senior Advisor, Pacific Business Group on Health)
-Fred Martin Jacobs, MD, JD, FACP, FCCP, FCLM (Executive Vice President and Director, Saint Barnabas Quality Institute)
-Natalie Napolitano, MPH, RRT-NPS (Respiratory Therapist, Inova Fairfax Hospital)
-Russell Robbins, MD, MBA (Principal and Senior Clinical Consultant, Mercer)

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The risk-standardized mortality rates obtained from the COPD mortality measure as specified will provide an accurate reflection of quality."

On a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5=Moderately Agree, and 6=Strongly Agree), 10 of 11 TEP members responded to the survey question as follows: Strongly Disagreed (1), Somewhat Agreed (3), Moderately Agreed (4), and Strongly Agreed (2). Of the TEP members who responded, 90% agreed (60% moderately or strongly agreed) that the measure will provide an accurate reflection of quality. We therefore gave the measure a moderate rating for face validity. In summary, these results demonstrated TEP agreement with the overall face validity of the measure as specified.

References:

Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS One 2011; 6(4):e17401.

Keenan PS, Normand SL, Lin Z, et al. 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 2008; 1(1):29-37.

Krumholz HM, Brindis RG, Brush JE, et al. 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. January 24, 2006; 113(3):456-462. [1]

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation 2006; 113(13):1683-92. [2]

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation 2006; 113:1693-1701. [3]

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Shahian DM, He X, O'Brien S, et al. Development of a Clinical Registry-Based 30-Day Readmission Measure for Coronary Artery Bypass Grafting Surgery. Circulation 2014; DOI: 0.1161/CIRCULATIONAHA.113.007541. Published online before print June 10, 2014.

Suter L, Wang C, Araas M, et al. Hospital-Level 30-Day All-Cause Unplanned Readmission Following Coronary Artery Bypass Graft Surgery (CABG): Updated Measure Methodology Report. 2014; http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890352615&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmsn_CABG_MeasMethd_Rpt_060314.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed November 4, 2015.

ICD-9 to ICD-10 Conversion
Statement of Intent
[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion
ICD-10 codes were identified using 2015 GEM mapping software. We then 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. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The performance of the first half of the split sample (development sample) and second half of the split sample (validation sample) from **Dataset 2** was similar. The areas under the receiver operating characteristic (ROC) curve for the two models are 0.720 and 0.723, respectively.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The results between the first half of the split sample and second half of the split sample from **Dataset 2** proved to be similar in each of the model testing that was performed. The ROC results were nearly identical and in line with other mortality models.

Validity as Assessed by External Groups:
The face validity testing results demonstrated TEP agreement with overall face validity of the measure as specified.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In **Dataset 1**:

Exclusion	**N**	**%**	**Distribution across hospitals (N=3,751): Min, 25th, 50th,**

			75th percentile, max
1. Inconsistent or unknown vital status or other unreliable demographic data	12	<0.01%	(0.00, 0.00, 0.00, 0.00, 0.96)
2. Enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission	14,136	1.37%	(0.00, 0.00, 0.71, 1.83, 29.41)
3. Discharged against medical advice (AMA)	6,020	0.58%	(0.00, 0.00, 0.00, 0.69, 20.00)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusion 1 (patients with inconsistent or unknown vital status or other unreliable demographic [age and gender] data), we do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive. This exclusion accounts for <0.01% of all index admissions excluded from the initial index cohort.

Exclusion 2 (patients enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission), these patients are likely continuing to seek comfort measures only; mortality is not necessarily an adverse outcome or signal of poor quality care. This exclusion accounts for 1.37% of all index admissions excluded from the initial index cohort.

Exclusion 3 (patients who are discharged AMA) accounts for 0.58% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 42 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Our approach to risk adjustment was 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 measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al., 2007). At the patient level, each model adjusts the log-odds of mortality within 30-days of admission for age, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital-specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. 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.

Clinical Factors
Candidate and Final Risk-adjustment Variables: The original measure was developed using Medicare FFS claims data. Candidate variables were patient-level risk-adjustors that are expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusted for case differences based on the clinical status of the patient at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission. In addition, only comorbidities that conveyed information about the patient at that time or in the 12-months prior, and not complications that arose during the course of the hospitalization were included in the risk-adjustment.

The final set of risk-adjustment variables is:
• Age-65 (years, continuous) for patients aged 65 or over cohorts; or Age (years, continuous) for patients aged 18 and over cohorts
- History of mechanical ventilation
- Sleep apnea
- Respirator dependence/respiratory failure
- Cardio-respiratory failure or shock
- Congestive heart failure

- Coronary atherosclerosis or angina
- Specified arrhythmias and other heart rhythm disorders
- Vascular or circulatory disease
- Fibrosis of lung or other chronic lung disorders
- Asthma
- Pneumonia
- Pleural effusion/pneumothorax
- Other lung disorders
- Metastatic cancer or acute leukemia
- Lung, upper digestive tract, and other severe cancers
- Lymphatic, head and neck, brain, and other major cancers; breast, prostate, colorectal and other cancers and tumors; other respiratory and heart neoplasms
- Other digestive and urinary neoplasms
- Diabetes mellitus (DM) or DM complications
- Protein-calorie malnutrition
- Disorders of fluid/electrolyte/acid-base
- Other endocrine/metabolic/nutritional disorders
- Other gastrointestinal disorders
- Osteoarthritis of hip or knee
- Other musculoskeletal and connective tissue disorders
- Iron deficiency or other unspecified anemias and blood disease
- Dementia or other specified brain disorders
- Drug/alcohol abuse, without dependence
- Other psychiatric disorders
- Hemiplegia, paraplegia, paralysis, functional disability
- Mononeuropathy, other neurological conditions/injuries
- Hypertension and hypertensive disease
- Stroke
- Retinal disorders, except detachment and vascular retinopathies
- Other eye disorders
- Other ear, nose, throat and mouth disorders
- Renal failure
- Decubitus ulcer or chronic skin ulcer
- Other dermatological disorders
- Trauma
- Vertebral fractures
- Major complications of medical care and trauma

Socioeconomic Status (SES) Factors and Race

We selected variables representing socioeconomic status (SES) factors and race for examination based on a review of literature, conceptual pathways, and feasibility. In section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SES and race may influence 30-day mortality.

Our conceptualization of the pathways by which patient SES or race affects 30-day mortality is informed by the literature.

Literature Review of Socioeconomic Status (SES) and Race Variables and COPD Mortality

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following COPD hospitalization, a literature search was performed with the following exclusion criteria: international studies, articles published more than 10 years ago, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on SES or race and COPD mortality. Twenty-eight studies were reviewed by title and abstract, and twenty-seven studies were excluded from full-text review. While limited data were identified meeting these criteria, the study reviewed found that health disparities indicators were associated with increased risk of COPD mortality (Lewis et al., 2009).

Causal Pathways for Socioeconomic Status (SES) and Race Variable Selection

Although some recent literature evaluates the relationship between patient SES or race and the mortality outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with mortality. The SES factors that have been examined in the mortality literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients, and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Alter et al., 2014; Taksler et al., 2012). Neighborhood/community-level variables use information from sources such as the American Community Survey (ACS) as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the Agency for Healthcare Research and Quality (AHRQ)-validated SES index score (Blum et al., 2014). Hospital-level variables measure attributes of the hospital which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital.

The conceptual relationship, or potential causal pathways by which these possible SES risk factors influence the risk of mortality following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider.

1. **Relationship of socioeconomic status (SES) factors or race to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These SES risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job, lack of childcare), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment**.**

In addition to SES risk factors, studies have shown that worse health status is more prevalent among African-American patients compared with white patients. The association between race and worse health is in part mediated by the association between race and SES risk factors such as poverty or disparate access to care associated with poverty or neighborhood. The association is also mediated through bias in healthcare as well as other facets of society.

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus patients with low income are more likely to be seen in lower quality hospitals, which can contribute to increased risk of mortality following hospitalization

(Jha et al., 2011; Reames et al., 2014). Similarly African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner et al., 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to mortality risk is that patients may not receive equivalent care within a facility. For example, African-American patients have been shown to experience differential, lower quality, or discriminatory care within a given facility (Trivedi et al., 2014). Alternatively, patients with SES risk factors such as lower education may require differentiated care – e.g. provision of lower literacy information – that they do not receive.

4. **Influence of socioeconomic status (SES) on mortality risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of mortality without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are complex to distinguish analytically. They also have different implications on the decision to risk adjust or not. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in 1.8, the following SES and race variables were considered:

- African American race (as compared to all others)
- Dual eligible status
- AHRQ SES index score

We assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

References:

Alter DA, Franklin B, Ko DT, et al. Socioeconomic status, functional recovery, and long-term mortality among patients surviving acute myocardial infarction. PloS one 2014; 8:e65130.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation Cardiovascular quality and outcomes 2014; 7:391-7.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs 2011; 30:1904-11.

Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia. 2006. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 19, 2015.

Lewis DR, Clegg LX, Johnson NJ. Lung disease mortality in the United States: the National Longitudinal Mortality Study. Int J Tuberc Lung Dis 2009 Aug; 13(8):1008–1014.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

Taksler GB, Keating NL, Cutler DM. Explaining racial differences in prostate cancer mortality. Cancer 2012; 118:4280-9.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Below is a table showing the final variables in the model with associated odds ratios.

Final Model Variables (variables meeting criteria in field 2b4.3) (**Dataset 1**)

Variable	07/2011-06/2014 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.04 (1.04 -1.04)
Sleep apnea (ICD-9 codes 327.20, 327.21, 327.23, 327.27, 327.29,780.51, 780.53, 780.57)	0.92 (0.89 -0.94)
History of mechanical ventilation (ICD-9 codes 93.90, 96.70, 96.71,96.72)	1.28 (1.24 -1.31)
Respirator dependence/respiratory failure (CC 77-78)	0.87 (0.81 -0.93)
Cardio-respiratory failure or shock (CC 79)	1.45 (1.42 -1.48)
Congestive heart failure (CC 80)	1.26 (1.23 -1.28)
Coronary atherosclerosis or angina (CC 83-84)	0.96 (0.94 -0.98)
Specified arrhythmias and other heart rhythm disorders (CC 92-93)	1.07 (1.05 -1.09)
Vascular or circulatory disease (CC 104-106)	1.04 (1.02 -1.06)
Fibrosis of lung or other chronic lung disorders (CC 109)	1.13 (1.11 -1.16)
Asthma (CC 110)	0.69 (0.67 -0.71)
Pneumonia (CC 111-113)	1.25 (1.23 -1.27)
Pleural effusion/pneumothorax (CC 114)	1.18 (1.15 -1.21)
Other lung disorders (CC 115)	0.84 (0.83 -0.86)
Metastatic cancer or acute leukemia (CC 7)	2.37 (2.28 -2.47)
Lung, upper digestive tract, and other severe cancers (CC 8)	1.83 (1.78 -1.89)
Lymphatic, head and neck, brain, and other major cancers; breast, prostate, colorectal and other cancers and tumors; other respiratory and heart neoplasms (CC 9-11)	1.01 (0.99 -1.04)
Other digestive and urinary neoplasms (CC 12)	0.82 (0.79 -0.85)
Diabetes mellitus (DM) or DM complications (CC 15-20, 119-120)	0.96 (0.94 -0.97)
Protein-calorie malnutrition (CC 21)	2.12 (2.07 -2.17)
Disorders of fluid/electrolyte/acid-base (CC 22-23)	1.13 (1.11 -1.16)
Other endocrine/metabolic/nutritional disorders (CC 24)	0.82 (0.81 -0.84)
Other gastrointestinal disorders (CC 36)	0.85 (0.84 -0.87)
Osteoarthritis of hip or knee (CC 40)	0.74 (0.71 -0.76)
Other musculoskeletal and connective tissue disorders (CC 43)	0.84 (0.83 -0.86)
Iron deficiency or other unspecified anemias and blood disease (CC 47)	1.28 (1.25 -1.30)
Dementia or other specified brain disorders (CC 49-50)	1.18 (1.16 -1.21)
Drug/alcohol abuse, without dependence (CC 53)	0.88 (0.86 -0.90)
Other psychiatric disorders (CC 60)	1.15 (1.13 -1.17)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.03 (0.99 -1.07)
Mononeuropathy, other neurological conditions/injuries (CC 76)	0.87 (0.85 -0.89)
Hypertension and hypertensive disease (CC 90-91)	0.84 (0.82 -0.86)
Stroke (CC 95-96)	0.94 (0.91 -0.97)

Variable	07/2011-06/2014 OR (95% CI)
Retinal disorders, except detachment and vascular retinopathies (CC 121)	0.93 (0.90 -0.95)
Other eye disorders (CC 124)	0.89 (0.87 -0.91)
Other ear, nose, throat and mouth disorders (CC 127)	0.80 (0.79 -0.82)
Renal failure (CC 131)	1.08 (1.06 -1.10)
Decubitus ulcer or chronic skin ulcer (CC 148-149)	1.35 (1.31 -1.39)
Other dermatological disorders (CC 153)	0.91 (0.89 -0.92)
Trauma (CC 154-156, 158-161)	1.03 (1.01 -1.06)
Vertebral fractures (CC 157)	1.29 (1.25 -1.34)
Major complications of medical care and trauma (CC 164)	0.84 (0.81 -0.87)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Variation in prevalence of the factor across measured entities

The prevalence of SES factors and African-American patients in the COPD cohort varies across measured entities. The median percentage of dual eligible patients is 19.1% (interquartile range [IQR] 12.9%-26.8%). The median percentage of African-American patients is 2.4% (IQR 0.0%-9.1%). The median percentage of patients with an AHRQ SES index score equal to or below 45.1 is 15.2% (IQR 2.9%-44.7%).

Empirical association with the outcome (univariate)

The patient-level observed COPD mortality rate is lower for dual eligible patients, 7.2%, compared with 7.8% for all other patients. The mortality rate for African-American patients was also lower at 5.7% compared with 7.9% for patients of all other races. Similarly the mortality rate for patients with an AHRQ SES index score equal to or below 45.1 was 7.1% compared with 7.9% for patients with an AHRQ SES index score above 45.1.

Incremental effect of SES variables and race in a multivariable model

We then examined the strength and significance of the SES variables and race in the context of a multivariable model. Consistent with the above findings, when we include any of these variables in a multivariate model that includes all of the claims-based clinical variables, the effect size of each of these variables is small and protective. We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model. Furthermore we find that the addition of any of these variables into the model has little to no effect on hospital performance. We examined the change in hospitals' RSMRs with the addition of any of these variables. The mean absolute change in hospitals' RSMRs when adding a dual eligibility indicator is -0.0004% with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.99989. The mean absolute change in hospitals' RSMRs when adding a race indicator is -0.0025% with a correlation coefficient between RSMRs for each hospital with and without race added of 0.99348. The mean absolute change in hospitals' RSMRs when adding an indicator for a low AHRQ SES index score is -0.0015% with a correlation coefficient between RSMRs for each hospital with and without an indicator for a low AHRQ SES index score added of 0.99423.

Overall, we find that among the SES and race variables that could be feasibly incorporated into this model the relationship between African-American race, dual-eligible status, and patients in the lowest quartile by AHRQ SES index score and mortality is in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also find that the impact of any of these indicators is very small to negligible on model performance and hospital profiling. Given the controversial nature of incorporating such variables into a risk-model we do not support doing so in a case that is unlikely to affect hospital profiling. Given these findings and the complex pathways that could explain any relationship between SES or race with mortality, which do not all support risk-adjustment, we did not incorporate SES variables and race into the measure.

We do think further investigation of lower mortality among dual eligible, African-American patients, and patients with a low AHRQ SES index score will be valuable and may shed light on proposed future modifications to the measure. Future reevaluation efforts will explore the relationship between SES and COPD once ICD-10 data are available.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance (**Dataset 1** & **Dataset 2**)
We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the COPD cohort:

Discrimination Statistics
(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome.
(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

Calibration Statistics
(3) 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.)

We tested the performance of the model for **Dataset 1** and **Dataset 2** described in section 1.7.

References:
Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* **17** (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

For the original measure development cohort (**Dataset 2**) the results are summarized below:
- 1st half of randomly split sample (development sample): C-statistic = 0.720; Dataset Predictive ability (lowest decile %, highest decile %) = (1.52, 23.74)
- 2nd half of randomly split sample (validation sample): C-statistic = 0.723; Dataset Predictive ability (lowest decile %, highest decile %) = (1.60, 23.78)

For the current measure cohort (version 4.0) (**Dataset 1**) the results are summarized below:
- C-statistic = 0.72
- Predictive ability (lowest decile %, highest decile %) = (1.4, 21.4)

For comparison of model with and without inclusion of SES factors, see above section.

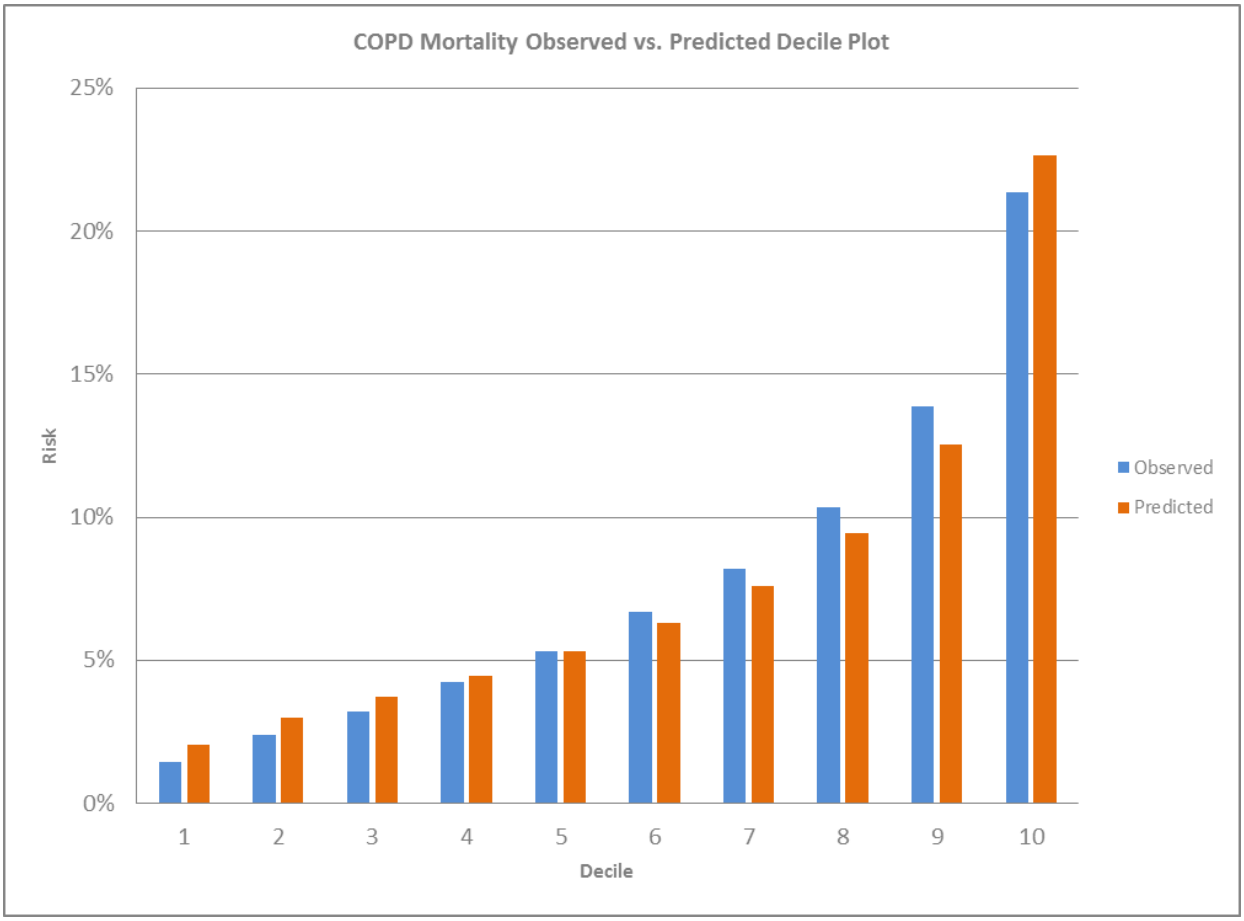
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the original measure development cohort (**Dataset 2**) the results are summarized below:

1st half of split sample (development sample): Calibration: (-0.034, 0.985)
2nd half of split sample (validation sample): Calibration: (0.009, 1.004)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2011 to June 2014 (**Dataset 1**).



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics

The c-statistics of 0.721 indicate good model discrimination (**Dataset 1**). The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Risk Decile Plots
Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates good discrimination of the model and good predictive ability.

Overall Interpretation
Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSMRs accounting for differences in hospital case-mix.

Application to Patients Aged 18 and Older (**Dataset 3**)
We applied the model to all-payer data from California. The analytic sample included 39,232 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.

To help determine whether the measure could be applied to an 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 R2) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.744). 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 (0.747 vs. 0.744); and (c) hospital-level risk-standardized rates were highly correlated (ICC = 0.999). Thus, the inclusion of the interactions did not substantively affect either patient-level model performance or hospital-level results.

We conducted this testing prior to specifying the measure for patients age 40 and over. Restricting the patient cohort to age 40 and over, however, is not likely to affect the results given that only 1.5% of patients were between the ages of 18 and 39. Therefore, the measure can be applied to all-payer data for patients 40 and older.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR 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 RSMR 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 cases in the three-year period.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from July 2011-June 2014 (**Dataset 1**), the median hospital RSMR was 7.7%, with a range of 4.8% to 13.8%. The interquartile range was 7.2%-8.2%.

Out of 4,658 hospitals in the U.S., 51 performed "better than the U.S. national rate," 3,611 performed "no different from the U.S. national rate," and 89 performed "worse than the U.S. national rate." 907 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for COPD that support measurement to reduce the variation.

Note: Over the three years of the measure reporting period, the COPD mortality rate has decreased from 7.7% (July 2011 to June 2012) to 7.4% (July 2013 to June 2014). Despite recent decreases in mortality rates nationally, the mortality rate for the 2015 public reporting period (July 2011 to June 2014) for COPD Medicare FFS patients is at 7.8%.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

N/A

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

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

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Administrative data are routinely collected as part of the billing process.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees associated with the use of this measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting
Program Name, Sponsor: Hospital Inpatient Quality Reporting (Hospital IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital IQR program was originally mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. Initially, the MMA provided for a 0.4 percentage point reduction in the annual market basket (the measure of inflation in costs of goods and services used by hospitals in treating Medicare patients) update for hospitals that did not successfully report. The Deficit Reduction Act of 2005 increased that reduction to 2.0 percentage

points.

In addition to giving hospitals a financial incentive to report the quality of their services, the hospital reporting program provides CMS with data to help consumers make more informed decisions about their health care. Some of the hospital quality of care information gathered through the program is available to consumers on the Hospital Compare website at: www.hospitalcompare.hhs.gov.

Geographic area and number and percentage of accountable entities and patients included:
The Hospital IQR program includes all IPPS non-federal acute care hospitals. The number and percentage of accountable hospitals included in the program, as well as the number of patients included in the measure, varies by reporting year. For 2015 public reporting, the RSMR was reported for 4,658 hospitals across the U.S. The final index cohort includes 780,879 admissions.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A. This measure is currently publicly reported.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A. This measure is currently publicly reported.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

There has been significant progress in 30-day RSMR for COPD. The median 30-day RSMR decreased by 0.2 absolute percentage points from July 2011-June 2012 (median RSMR: 7.6%) to July 2013-June 2014 (median RSMR: 7.4%). The median hospital RSMR from July 2011-June 2014 was 7.7% (IQR 7.2% - 8.2%). In addition, hospitals with a high proportion of dual eligible and African American patients achieve a similar range of performance as compared with hospitals with a low proportion of these patients. In addition, hospitals with a low proportion of patients with AHRQ SES index scores equal to or below 45.1 perform similarly to hospitals with a high proportion of patients with AHRQ SES index scores equal to or below 45.1. These results indicate that both groups of hospitals can perform well on the measure.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure development, model testing, or re-specification. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0275 : Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)
0700 : Health-related Quality of Life in COPD patients before and after Pulmonary Rehabilitation
0701 : Functional Capacity in COPD patients before and after Pulmonary Rehabilitation
1891 : Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses 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.**
We did not include in our list of related measures any non-outcome (for example, process) measures with the same target population as our measure. Our measure cohort was heavily vetted by clinical experts, a technical expert panel, and a public comment period. Additionally, the measure, with the specified cohort, has been publicly reported since December 2014. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or

methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)
Co.4 **Point of Contact:** Karen, Dorsey, karen.dorsey@yale.edu, 203-764-5700-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
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.

Our measure development team consisted of the following members:
Laura M. Grosso, PhD, MPH
Peter Lindenauer, MD, MSc
Changqin Wang, MD, MS
Shantal Savage, BA
Jaymie Potteiger, MPH
Yun Wang, PhD
Zameer Abedin, BA
Lori L. Geary, MPH
Elizabeth E. Drye, MD, SM
Harlan M. Krumholz, MD, SM

Technical Expert Panel Members:
Darlene Bainbridge, RN, MS, NHA, CPHQ, CPHRM President/CEO, Darlene D. Bainbridge & Associates, Inc.

Robert A. Balk, MD, Director of Pulmonary and Critical Care Medicine, Rush University Medical Center

Dale Bratzler, DO, MPH, President and CEO, Oklahoma Foundation for Medical Quality

Scott Cerreta, RRT, Director of Education, COPD Foundation

Gerard J. Criner, MD, Director of Temple Lung Center and Divisions of Pulmonary and Critical Care Medicine, Temple University

Guy D'Andrea, MBA, President, Discern Consulting

Jonathan Fine, MD, Director of Pulmonary Fellowship, Research and Medical Education, Norwalk Hospital

David Hopkins, MS, PhD, Senior Advisor, Pacific Business Group on Health

Fred Martin Jacobs, MD, JD, FACP, FCCP, FCLM, Executive Vice President and Director, Saint Barnabas Quality Institute

Natalie Napolitano, MPH, RRT-NPS, Respiratory Therapist, Inova Fairfax Hospital

Russell Robbins, MD, MBA, Principal and Senior Clinical Consultant, Mercer

Working Group Panel Members:
David Au, MD, MS, Investigator, VA Puget Sound Healthcare System, Northwest HSR&D Center of Excellence; Associate Professor of Medicine, Department of Medicine, University of Washington

Jerry Krishnan, MD, PhD, Associate Professor, Departments of Medicine and Health Studies, University of Chicago; Director, Asthma Center and Refractory Obstructive Lung Disorders Clinic, University of Chicago

Todd Lee, PharmD, PhD, Associate Professor, Departments of Pharmacy Practice and Pharmacy Administration, University of Illinois at Chicago; Senior Investigator, Center for Management of Complex Chronic Care (CMC3), Hines VA Hospital

Richard Mularski, MD, MCR, MSHS, Clinical Investigator, Center for Health Research, Kaiser Permanente; Clinical Assistant Professor of Medicine, Oregon Health & Science University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2014

Ad.3 **Month and Year of most recent revision:** 07, 2015

Ad.4 **What is your frequency for review/update of this measure?** Annual

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.6 **Copyright statement:** N/A

Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** N/A

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2794
Measure Title: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: A PQMP Measure
Measure Steward: University Hospitals Cleveland Medical Center
Brief Description of Measure: This measure estimates the rate of emergency department visits for children ages 2 – 21 who are being managed for identifiable asthma. The measure is reported in visits per 100 child-years.
Developer Rationale: ED visits for children with asthma is an outcome measure of intrinsic value. It represents utilization of an expensive service and constitutes a burden on children and their families. There is abundant evidence that ED visits are common, may be reduced through improved primary care or community-based interventions, and demonstrate disparities (1-11, 12-19). Asthma is generally recognized to be an ambulatory care sensitive condition. Nonetheless, we perceive and our panel articulated that the rate for ED visits ought not to be 0. So while in general a lower rate represents preferable care, too low a rate could indicate insufficient access to emergency room services. Our overarching conceptual framework that extends beyond this measure is shown in the evidence form.

Our measure benefits from a formal development process, CAPQuaM's 360 degree method, which is described in more detail in the measure testing form. The measure and its specifications result from a formal development process for this measure incorporated stakeholder input including a parent focus group, meeting with The Mount Sinai Pediatrics Department's Parent Advisory Council, interviews with primary care clinicians and ED physicians, the CAPQuaM's multidisciplinary scientific team, which includes investigators, a steering committee and a senior advisory board of nationally prominent figures. The measure also benefits from a national multidisciplinary Expert Panel which utilized a RAND type modified Delphi method to guide our specifications.

When epidemiologists describe how frequently something occurs the preferred measure is typically an incidence density, or rate. In contrast to a risk or proportion, the incidence density has as its denominator a measure of the extent of potential exposure in the population, expressed in people-years. This measure represents an advance in the measurement of healthcare performance for children: it incorporates this formulation both to enhance its interpretation (because it has a specific epidemiological meaning) and to limit distortion if sick children move in or out of eligibility for the measure. (20) Further clinical evidence of gaps are demonstrated in the description by NHLBI's NAELPP guideline, cited in the evidence form, Schatz and colleagues study describing the relationship between asthma control and asthma exacerbations in managed care (21), and Fuhlbrigge et al's confirmation that medications can work to reduce ED visits for asthma but are used sub optimally (22).

References

1. Adams, R.J., B.J. Smith, and R.E. Ruffin, Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. Thorax, 2000. 55(7): p. 566-73.
2. Bahadori, K., et al., Economic burden of asthma: a systematic review. BMC Pulm Med, 2009. 9: p. 24.
3. Weiss, K.B., P.J. Gergen, and T.A. Hodgson, An economic evaluation of asthma in the United States. N Engl J Med,

1992. 326(13): p. 862-6.
4. Coventry, J.A., M.S. Weston, and P.M. Collins, Emergency room encounters of pediatric patients with asthma: cost comparisons with other treatment settings. JAmbul Care Manage, 1996. 19(2): p. 9-21.
5. Association, A.L. American Lung Association Fact Sheet on Asthma and Children. October 2012 [cited 2013 May 20]; Available from: http://www.lung.org/lung-disease/asthma/resources/facts-and-figures/asthmachildren-fact-sheet.html#5.
6. Fuhrman, C., et al., Hospitalizations for asthma in children are linked to undertreatment and insufficient asthma education. J Asthma, 2011. 48(6): p. 565-71.
7. Sawicki, G.S., et al., Uncontrolled asthma in a commercially insured population from 2002 to 2007: trends, predictors, and costs. J Asthma, 2010. 47(5): p. 574-80.
8. Manice, M., Exploring the relationship between parental shared decision-making practices and acute asthma exacerbations among children age 0-17. 2013, Icahn School of Medicine at Mount Sinai: New York, NY.
9. Cerdan, N.S., et al., Asthma severity in children and the quality of life of their parents. Appl Nurs Res, 2012. 25(3): p. 131-7.
10. Fiese, B., et al., Family climate of routine asthma care: associating perceived burden and mother-child interaction patterns to child well-being. Fam Process, 2008. 47(1): p. 63-79.
11. Okelo, S.O., et al., Emotional quality-of-life and outcomes in adolescents with asthma. J Pediatr, 2004. 145(4): p. 523-9.
12. Talreja, N., et al., Modifiable factors associated with severe asthma exacerbations in urban patients. Ann Allergy Asthma Immunol, 2012. 109(2): p.128-32.
13. Auger, K.A., et al., Medical home quality and readmission risk for children hospitalized with asthma exacerbations. Pediatrics, 2013. 131(1): p. 64-70.
14. Ducharme, F.M., et al., Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control. Am J Respir Crit Care Med, 2011.183(2): p. 195-203.
15. Farber, H.J., Optimizing maintenance therapy in pediatric asthma. Curr Opin Pulm Med, 2010. 16(1): p. 25-30.
16. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, Relationship between pediatric primary provider visits and acute asthma ED visits. Pediatr Pulmonol, 2007.42(11): p. 1041-7.
17. Self, T.H., et al., Reducing emergency department visits and hospitalizations in African American and Hispanic patients with asthma: a 15-year review. J Asthma, 2005. 42(10): p. 807-12.
18. Lara, M., et al., Reducing quality-of-care disparities in childhood asthma: La Red de Asma Infantil intervention in San Juan, Puerto Rico. Pediatrics, 2013. 131 Suppl 1: p. S26-37.
19. Oraka, E., et al., Racial and ethnic disparities in current asthma and emergency department visits: findings from the national health interview survey, 2001-2010. J Asthma, 2013. 50(5): p. 488-96.
20. Rothman, K.J., Modern Epidemiology. 3rd edition ed. 2008, Philadelphia, PA: Lippincott, Williams and Wilkins.
21. Schatz M, Zeiger RS, Yang ST, et al. "Relationship of asthma control to asthma exacerbations using surrogate markers within a managed care database." Am J Managed Care. 16(5):327-333, 2010.
22. Fuhlbrigge A, Carey VJ, Adams RJ. "Evaluation of asthma prescription measures and health system performance based on emergency department utilization." Med Care 42(5):1-7, 2004.

Numerator Statement: The numerator uses the number of undesirable utilization outcomes (i.e., claims for ED visits or hospitalizations for asthma) experienced by children who are managed for identifiable asthma to estimate the number of emergency room visits

Denominator Statement: The denominator represents the person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years.

Denominator Exclusions: Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 Code: 496), Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03, 277.09), or Emphysema diagnosis (ICD-9 code 492xx).

These exclusion incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting

month and for the reporting month (a total of three consecutive months ending in the reporting month).

Measure Type: Outcome
Data Source: Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
Level of Analysis: Health Plan, Integrated Delivery System, Population : Community, Population : County or City, Population : National, Population : Regional, Population : State

New Measure -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

The developer provides the following rationale for this outcome measure:

1. Accessible, high-quality primary care reduces the need for emergency department (ED) visits by decreasing the number of children who have acute breakthrough episodes requiring the ED.
2. Accessible, high-quality primary care reduces the need for ED visits by decreasing the number of children who come to the ED for asthma care better performed in the office setting.

- A systematic review of the body of evidence is not required for outcome measures.
- The evidence for this measure is based on clinical practice guidelines for asthma control from the National Heart and Lung and Blood Institutes (NHLBI), dated 2006.
- Although not required per NQF guidance, the developer conducted a literature review (September 2004 to March 2006) on asthma care: A total of 4,747 abstracts and articles were reviewed.

Question for the Committee:

- *Is there at least one thing that the measured entity can do to achieve a change in the measure results? (The developer has indicated the following levels of analyses: Health Plan; Integrated Delivery System; Population: Community, County or City, Regional, State, or National.)*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer reports:

- There is abundant evidence that ED visits for asthma are common, may be reduced through improved primary care or community-based interventions, and demonstrate disparities.
- NHLBI NAELPP guideline provides a description of clinical evidence of gaps.
- The developer reports overall rate of ED visits for asthma in NY State Medicaid Managed Care in 2012 is 28.95 per 100 child-years in children 2 to 18 years and 29.25 for children 2 to 21 years.
    - By age stratum the rates are 47.4 visits per 100 child-years for children 2 to 4 years, 26.0 for children 5 to 11 years, 22.7 for adolescents 12 to 18 years, and 34.1 for adolescents 19 to 21

years.
    - The developer provides additional data demonstrating expected seasonal variations in performance rates.

Disparities

The developer reports:

- Asthma is a critical problem with racial and ethnic disparities and varies by urbanicity. The developer's analysis of National Survey of Children's Health data (NSCH, 2011/12), estimates that 10.3 million children in the United States have been told that they have asthma. Of these children, 7.6 million live in more urban areas that are characterized as metropolitan statistical areas (MSAs), with an asthma prevalence rate of 15.4%.
- The developer reports that, on a yearly and a monthly basis, differences exist in performance by age, urbanicity, race/ethnicity, and level of poverty. Additionally, it identifies disparities in cross tabulations—e.g., the performance rate for children 2 to 4 years in large metropolitan areas is 52.6 visits per 100 child-years compared to those in small metropolitan areas with 26.2 visits per child year, in micropolitan areas with 18.3 visits/100 child-years, and in rural areas with 12.3 visits per 100 child-years.
- The developer reports racial and ethnic differences were notable:
    - For children ages 2 to 4 years, the rate in non-Hispanic Whites was 18.4 visits per 100 child-years, in Asians 19.3 visits per 100 child-years, in Hispanics 53.9 child-years, and in non-Hispanic Blacks 74 visits per 100 child-years.
    - The disparities regardless of age were Black, 41.99 visits per 100 child-years; White, 14.79 visits per 100 child-years; Hispanic, 31.91 visit per 100 child-years.

Questions for the Committee:

- *Does the Committee believe there is a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The evidence reflects that the measure supports that a change in process including access to primary care providers may improve management of asthma and result in a reduction in ED visits for acute break through episodes and overall asthma ED visits due to better asthma care management in an outpatient setting. This is an outcomes measure that uses ED visits as a proxy or representation of the care provided for asthma in a primary care provider's office. The desired outcome is better asthma care/management in a PCP setting resulting in reduced ED visits.

**Measure could lead to meaningful improvement

**Partially - so a qualified yes. ED utilization affected by this but also other factors.

1b. Performance Gap

Comments:

**The developer relied on a literature search primarily for its evidence on optimal performance and to demonstrate gaps in care. The measure description also notes differences that exist in performance by age, urbanicity, race/ethnicity, and level of poverty. They also cited from a literature review that there were gaps in outcomes between large and small urban areas. The developer noted significant racial disparity gaps. All data was obtained through literature review to develop the measure.

**Certainly there is likely a gap in care.

**Not compelling - variation exists, but how much by primary care differences or quality vs other factors - esp. socioeconomic and geographic - is not clear.

1c. High Priority (previously referred to as High Impact)
Comments:
**Not applicable.
**na

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

Specifications:

- The developer defines the numerator as: *The number of undesirable utilization outcomes (i.e., claims for ED visits or hospitalizations for asthma) experienced by children who are managed for identifiable asthma to estimate the number of emergency room visits.*
- The denominator for this measure is: *The person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years.*
- The general data elements are: age, race and ethnicity; insurance type (Medicaid, Private, Uninsured); benefit type among insured (HMO, PPO, FFS, Medicaid Primary Care Case Management Plan [PCCM], Other); ZIP code or State and County of residence (and FIPS where available).
- The administrative data with billing and diagnosis codes are: asthma-related visits to an emergency department, or hospitalization; asthma medication prescriptions; insurance benefit type; ZIP code or State and County of residence (and FIPS where available); face and ethnicity (from hospital administrative data or charts if not in administrative data from plan).
    - The developer states "pharmacy data are not critical," and notes that if pharmacy data are not available the measure should be reported with notation that pharmacy data were not used for the assessment of eligibility.
- The numerator and denominator details include the CPT and ICD-9 codes; ICD-10 codes are included in an attachment.
- This outcome measure is not risk adjusted.
- The calculation algorithm is stated in S.18.

Questions for the Committee:

- *Are the definitions and codes for "managed for identifiable asthma" and "asthma related medication" appropriate? Are they specific enough so they can be reliably collected by different parties?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*
- *Is the potential variability in access to/inclusion of pharmacy a concern?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developer relied on pre-existing data element-level validity testing to identify children who are being managed for identifiable asthma. Per NQF guidance separate reliability testing is not required if data element-level validity testing is performed.

Results of reliability testing

- Not applicable; see discussion on validity testing at the data element level.

Guidance from the Reliability Algorithm : Not Applicable

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

- The goal of the measure is to assess how many children are visiting the ED for asthma treatment. According to the developer, ED visits for asthma are a function of a sick child that needs to be seen; poor access to high-quality primary care; or poor quality management of a chronic condition. The rate should be low, but not zero. The numerator of children with undesirable visits and a denominator of children with identified asthma are consistent with this evidence.

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer relies on literature to support its conclusion of the validity of administrative data elements to identify children who are being managed with identifiable asthma. Per NQF policy:
  - Validity testing at the data element level obviates the need for reliability testing at the data element level provided the critical data elements of the measure as specified are tested.
  - Prior evidence of validity of data elements can be used, including published data, provided it includes the same data elements; uses the same data type; and is conducted on an appropriate sample (i.e., representative, adequate numbers, etc.)
  - The developer attests that the data elements match those assessed in the literature.
- The developer also cites two NQF-endorsed measures from NCQA (*NQF 1799: Medication Management for People With Asthma* and *NQF 1800: Asthma Medication Ratio*, as well as *NQF 0036: Use of Appropriate Medications for People with Asthma*, which is no longer being maintained) as evidence of data-element level validity.
  - The developer acknowledges #1799 and #1800 are not directly applicable because they were tested at the score level. The developer notes, however, "the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to [reliably] produce valid scores."
  - The developer states there is "nearly complete overlap of the denominator codes for this measure and there is overlap of the denominator elements." The developer states that where codes differ, they were specific decisions by its expert panel.
- The developer cites face validity, but did not specifically assess face validity of at the computed measure score level, as required by NQF for face validity testing.
- The developer used NY State Medicaid Managed Care claims data for its analyses.

Validity testing results:

- As noted earlier, the developer refers to NCQA's data for three measures—#0036, #1799, and #1800—because these measures also rely on nearly complete overlap" in the codes necessary to identify the target population of children with identifiable asthma.
  - The developer cites these measures as evidence that administrative data have the capacity to appropriately identify patients with asthma because the measure scores distinguish signal from noise. That is, were these measures inadequately able to do so for their denominator populations, the empirical reliability testing would not have yielded high reliability scores.
- The developer provides information about various articles related to the use of administrative data for performance measurement.
- The developers interpret the measure to be a valid estimate of the rate of ED visits and an even better estimate of undesirable outcomes from asthma.

Questions for the Committee:

- *Do you agree that the score from this measure, as specified, is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following information:

- Denominator exclusions include: Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis, Cystic Fibrosis diagnosis, or Emphysema diagnosis. Children who were not

consecutively enrolled in the reporting plan for three consecutive months ending in the reporting month are excluded.
- There are no numerator exclusions.
- <=2.5% potentially eligible children were excluded by clinical diagnoses.
- Exclusions are clinical and represent construct validity rather than statistical considerations. Longer continuous enrollment requirements would harm the validity of the measure since more children with multiple diagnosis would have been excluded.

Questions for the Committee:
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☐ **Statistical model** ☒ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary
The developer provides the following information:
- "Specifications for this measure require stratification by age group and race/ethnicity." In follow-up to NQF staff, the developer clarified such stratification is "not for the purpose of risk adjustment, rather to demonstrate the presence or absence of disparities. This is fundamental to the measure, but does not replace [reporting] the topline rate."
- The developer notes additional stratification variables are optional (e.g., rurality/urbanicity and county level of poverty), but may be required by the accountability entity or reported by the reporting entity.
- The developer states biological data do not support risk adjustment to control for patient characteristics.
- The developer acknowledges the association of the risk factors with performance on the measure, but states risk adjustment is not justified by such differences as "either acceptable or unmodifiable by health care," and posits that evidence exists that primary care, adherence to guidelines, and other interventions can reduce or eliminate the impact of the risk factors.

Questions for the Committee:
- *The developer has provided data on differences in urbanicity, poverty, race and ethnicity, and SES (proxy of commercial vs. public insurance), but provides a rationale for not incorporating these variables in a risk model. Does the Committee concur with the developer that risk adjustment for SDS is not appropriate? The developer has indicated the following level of accountability/level of analysis: Health Plan; Integrated Delivery System; Population: Community, County or City, Regional, State, or National.)*
- *Is the Committee aware of evidence that contradicts the developer's rationale and analysis so that risk adjustment should be included?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:
- The developer used NY State Medicaid Managed Care claims data. The measure is specified at the health plan/integrated delivery system level and the population level, but not no data are presented that indicate meaningful difference among plans.
- The developer did analyze meaningful differences related to the stratification subpopulations (e.g., within a state), reporting it performed Chi-square analysis and t-statistics. The developer states "differences between major groups were statistically significant", p<0.05.
- The developer states the measure is sensitive enough to detect meaningful differences as observed

within a population, but no specific data are provided. The developer also notes:

- The sum of squares across populations is expected to be greater in distinct populations, so the developer expects the measure to perform well when comparing across populations as well.
- The effective sample size of within population comparisons (such as the developer conducted) is diminished by an intra-class correlation coefficient, so the developer expects greater power for equal sample size to detect differences between entities than it had for testing subpopulations within a single state.
- The developer states the signal to noise ratio is "very strong."

Questions for the Committee:

- *Does this measure identify meaningful differences about quality at the health plan level?*
- *Does this measure identify meaningful differences about quality at the population level?*

2b6. Comparability of data sources/methods:

- Not needed for this measure.

2b7. Missing Data

The developer notes the following:

- Since administrative claims are used, no missing data analyses were performed because the data set derived from sources that are contractually obligated to be provided to NYS Medicaid. The developer noted, however, when data "could not be obtained from any source":
  - If critical for calculation, delete patient from consideration for that report reporting month; if non-critical, include patient
  - The developer states "critical data include encounter data for the reporting month and some period of time in the assessment period." The developer also states race/ethnicity data are critical, and pharmacy data are not critical.
- The specifications note that pharmacy data may not be available, and any measure results presented without should indicate such. The developer reports that, based on alpha testing, the lack of such data would "reduce only slightly the number of children identified as having identifiable asthma." The developer cannot currently quantify the reduction in denominator, however, because NY State (the owner of the data used for these analyses) recently changed its data tables and have not yet had the capacity to complete the developer's request for such analyses.
  - The developer posits that systems unable to integrate pharmacy data into the eligibility analysis would have a "minimally higher risk population than those with pharmacy claims. The specifics of the definitions and the limited impact of pharmacy claims on eligibility combine to make the expected impact of this on the rate of ED visits to almost zero."
  - The developer states it included this component at the direction of its expert panel and, additionally, the NHLBI guideline states outcomes should not be adjusted for baseline risk. Based on this, the developer concludes "this does not truly disadvantage a reporting entity according to the guideline."

Questions for the Committee:

- *Is the Committee comfortable with the developer's designation of critical and non-critical data?*
- *For public reporting and accountability purposes, is the Committee concerned about the permitted variability with respect to pharmacy data's use in the specifications? Is the variability appropriate at the plan level? at the population level?*

Guidance from the Validity Algorithm: 1→2→4→5 (highest eligible rating is MODERATE)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The specifications are consistent with the evidence. The developer relies on literature to support its conclusion of the validity of administrative data elements to identify children who are being managed with identifiable asthma. The developer cites face validity, but did not specifically assess face validity of at the computed measure score level, as required by NQF for face validity testing.

**appropriate

**These seem valid, and the lack of pharmacy data for this outcome is not direct weakness - although it is a confounder as noted above (other factor besides primary care quality that impacts outcome.)

2a2. Reliability Testing

Comments:

**The developer did not do independent testing of the methodology or specifications. The developers interpret the measure to be a valid estimate of the rate of ED visits and an even better estimate of undesirable outcomes from asthma. The evidence cited support this conclusion.

**Children are likely to be able to be identified as noted

**The lack of explicit testing - esp. face validity - makes this measure unproven in my read. Extrapolating from other measures that target different populations for a different measure is not strong.

2b2. Validity Testing

Comments:

**Since the measure uses administrative data, missing data was not considered to be an issue to the validity of the measure. The numerator and denominator were clearly defined and utilized specified fields in administrative and/or electronic data.

**Should the pharmacy data be removed? Does it overly complicate the measure without reasonable improvement?

**I rate as moderate, with gaps re: risk adjustment (stratification alone will not achieve), and limited description of missing data impact or handling - it is possible missingness is a non-symmetric pattern, impacting the assessment of outcome.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:

**The developer relied on pre-existing data element-level validity testing to identify children who are being managed for identifiable asthma. Per NQF guidance separate reliability testing is not required if data element-level validity testing is performed.

**na

**N/A

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims.
- The developer reports there are no fees.

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**Data elements are routinely collected during service delivery. The elements are available in administrative data (claims/encounters) and would be available in an electronic health record. No concerns with the data collection strategy.
**reasonable
**No concerns - high level of feasibility

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
- The developer is working on specific plans for dissemination and use.
- The developer is discussing application and use of this measure with New York State Medicaid.
- The developer plans for the measure to be used for an accountability application within three years of NQF endorsement and public reporting within six years of initial endorsement.

Improvement results
- As a new measure, the developer does not present progress on improvement.
- The developer states a variety of stakeholders would benefit from this measure, e.g., clinicians, health systems, state and healthcare agencies, researches, etc.

Potential harms
- The developer reports no unintended negative consequences to individual or populations during testing.
- The developer reports possible unintended/negative consequences and recommends against the following:
    - Comparing individual health care professionals.
    - A single hospital comparison because this measure is intended to measure system performance not the hospital performance.
    - Measuring anything other than large practices or integrated delivery systems that own their own risk and manage inpatient and outpatient care or that have access to all payer data sources.

Feedback: No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal

program.

***Questions for the Committee**:*

- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**This measure is not currently being publicly reported. The results of the measure should result in opportunities to implement activities at the primary care provider level that should increase efficient and effective care and result in a reduction in ED visits related to asthma.
**would seem to be reasonable.
**Not in widespread use, so much opportunity; easily/already publicly available.
Limited information about specific provider impact programs that could follow, or unintended effects of the evaluation and scoring.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2852: Optimal Asthma Control
- 2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

Harmonization

- Measures have not been harmonized.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: CAPQuaM PQMP ASTHMA I: Rate of Emergency Department Visit Use for Children Managed for Persistent Asthma

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 10/2/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ **Health outcome:** ED asthma visits for children with identifiable asthma

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

The following diagram presents an overview of how CAPQuaM conceptualizes asthma ED visits for children with asthma.

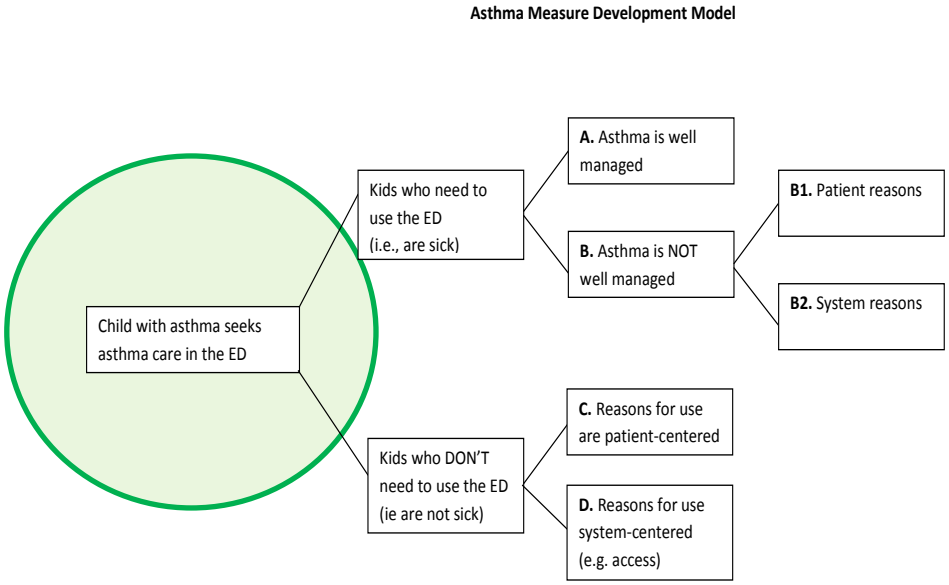


Figure Notes: The green circle highlights that this measure identifies which children who present to the emergency room should be considered to represent an ED visit for a child who is being managed for identifiable asthma.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

An abundant literature supports both that emergency department visits and hospitalizations are considered undesirable outcomes for asthma and that at a population level these undesirable outcomes can be reduced by better clinical management, including medication management, the use of asthma action plans, and effective and continuous primary care, among other things. Asthma is considered to be an ambulatory care sensitive condition further reinforcing the consensus in the field that utilization of ED visits and/or hospitalizations are generally (at the population level) preventable when managed in an ambulatory setting within our current knowledge.

3. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who have acute breakthrough episodes requiring the ED or inpatient setting.
4. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who come to the ED for asthma care better performed in the office setting.

As ED visits and/or hospitalizations can represent significant cost for families and for the system, asthma is the single most prevalent diagnosis leading to ED visits for children in the USA, urgent asthma visits to the ED can be disruptive for families, and both ED visits and hospitalizations are not free of iatrogenic and nosocomial risk, these outcomes have intrinsic importance.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*

NHLBI Asthma Guideline 2006 www.nhlbi.nih.gov/guidelines/asthma (NAEPP Guideline)

Quick Reference: Asthma control focuses on two domains:

1)reducing impairment --- the frequency and intensity of symptoms… and

2) reducing risk – the likelihood of future asthma attacks… [later described as "prevent exacerbations]

At the population level ED visits and hospitalizations represent failures of asthma control.

Systematic Review Okelo et al, Pediatrics 2013 132:5117-34

Demonstrates several tools are effective in enhancing the quality of care and reduce undesirable outcomes.

Cochrane Review: Chauhan et al Cochrane Database Syst Rev 201212:CD009611

Different approaches to treatment achieve different outcomes in children and adults (Daily achieves better asthma control than intermittent inhaled corticosteroids)

Systematic Review Mattke et al, Pediatrics 2009 123 S199-204

Identified multiple gaps in asthma care quality. Key outcomes identified include hospitalizations and emergency department visits. Identified large racial disparities in use of inhaled corticosteroids

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

From NHLBI:

http://www.nhlbi.nih.gov/health-pro/resources/lung/naci/asthma-info/asthma-guidelines.htm :

Asthma Guidelines

Following science-based guidelines *works*

Not only do they have the potential to improve a patient's *quality* of life; they can potentially *save a life*.

National asthma guidelines have been updated

In 2007, the National Asthma Education and Prevention Program (NAEPP), coordinated by the National Heart, Lung, and Blood Institute (NHLBI), released its third set of clinical practice guidelines for asthma. The Expert Panel Report 3—Guidelines for the Diagnosis and Management of Asthma (EPR-3) reflects the latest scientific advances in asthma drawn from a systematic review of the published medical literature by an NAEPP-convened expert panel. It describes a range of generally accepted best-practice approaches for making clinical decisions about asthma care.

The EPR-3 emphasizes the importance of asthma control and focuses on two domains—current impairment and future risk—by which to assess asthma severity (for initiating therapy) and asthma control (for ongoing monitoring). EPR-3 also includes an expanded section on childhood asthma (with an additional age group), new guidance on medications, new recommendations on patient education in settings beyond the physician's office, and new advice for controlling environmental exposures that can cause asthma symptoms.

Asthma can be controlled

Scientific evidence clearly shows that most people could control their asthma by following current asthma clinical practice guidelines. With proper care, people who have asthma can stay active, sleep through the night, and avoid having their lives disrupted by asthma attacks.

As a general rule, patients with well-controlled asthma should have:

- Few, if any, asthma symptoms.
- Few, if any, awakenings during the night caused by asthma symptoms.
- No need to take time off from school or work due to asthma.
- Few or no limits on full participation in physical activities.
- No emergency department visits.
- No hospital stays.
- Few or no side effects from asthma medicines.

NHLBI NAEPP Guideline

p. 36

KEYPOINTS: OVERVIEW OF MEASURES OF ASTHMA ASSESSMENT AND MONITORING

 The functions of assessment and monitoring are closely linked to the concepts of severity, control, and responsiveness to treatment:

— Severity: the intrinsic intensity of the disease process. Severity is measured most easily and directly in a patient not receiving long-term-control therapy.

— Control: the degree to which the manifestations of asthma (symptoms, functional impairments, and risks of untoward events) are minimized and the goals of therapy are met.

— Responsiveness: the ease with which asthma control is achieved by therapy.

 Both severity and control include the domains of current impairment and future risk:

— Impairment: frequency and intensity of symptoms and functional limitations the patient is experiencing or has recently experienced

— Risk: the likelihood of either asthma exacerbations, progressive decline in lung function (or, for children, reduced lung growth), or risk of adverse effects from medication

Page 37:

KEYDIFFERENCES FROM 1997 AND 2002 EXPERT PANEL REPORTS

 The key elements of assessment and monitoring are refined to include the separate, but related, concepts of severity, control, and responsiveness to treatment. Classifying severity is emphasized for initiating therapy; assessing control is emphasized for monitoring and adjusting therapy. Asthma severity and control are defined in terms of two domains: impairment and risk.

 The distinction between the domains of impairment and risk for assessing asthma severity and control emphasizes the need to consider separately asthma's effects on quality of life and functional capacity on an ongoing basis (i.e., in the present) and the risks it presents for adverse events in the future, such as exacerbations and progressive loss of pulmonary function. These domains of asthma may respond differentially to treatment.

… p.38

An important point linking asthma severity, control, and responsiveness is that the goals are

identical for all levels of baseline asthma severity. A patient who has severe persistent asthma

compared to a patient who has mild persistent asthma, or a patient who is less responsive to

therapy may require more intensive intervention to achieve well-controlled asthma; however, the

goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by

therapeutic intervention.

… page 41 regarding identification asthma, one key factor is:

The Expert Panel recommends that the clinician trying to establish a diagnosis of asthma

should determine that (EPR–2 1997):

 Episodic symptoms of airflow obstruction are present.

This is consistent with how we defined identifiable asthma…

Page 63

It is important to evaluate the frequency, rate of onset, severity, and causes of exacerbations…

severe exacerbations leading to ED visits and hospitalizations (Adams et al. 2000; Eisner et al.

2001; Ford et al. 2001; Lieu et al. 1998).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Methodology for report:

Overall Methods Used To Develop This Report

Background

In June 2004, the Science Base Committee of the NAEPP recommended to the NAEPP CC that its clinical practice guidelines for the diagnosis and management of asthma be updated. In September, under the leadership of Dr. Barbara Alving, M.D. (Chair of the NAEPP CC, and Acting Director of the NHLBI), a panel of experts was selected to update the clinical practice guidelines by using a systematic review of the scientific evidence for the treatment of asthma and consideration of literature on implementing the guidelines.

In October 2004, the Expert Panel assembled for its first meeting. Using EPR-2 1997 and EPR-Update 2002 as the framework, the Expert Panel organized the literature searches and subsequent report around the four essential components of asthma care, namely: (1) assessment and monitoring, (2) patient education, (3) control of factors contributing to asthma severity, and (4) pharmacologic treatment. Subtopics were developed for each of these four broad categories.

The steps used to develop this report include: (1) completing a comprehensive search of the literature; (2) conducting an indepth review of relevant abstracts and articles; (3) preparing evidence tables to assess the weight of current evidence with respect to past recommendations and new and unresolved issues; (4) conducting thoughtful discussion and interpretation of findings; (5) ranking strength of evidence underlying the current recommendations that are made; (6) updating text, tables, figures, and references of the existing guidelines with new findings from the evidence review; (7) circulating a draft of the updated guidelines through several layers of external review, as well as posting it on the NHLBI website for review and comment by the public and the NAEPP CC, and (8) preparing a final-report based on consideration of comments raised in the review cycle.

Systematic Evidence Review Overview

Inclusion/Exclusion Criteria

The literature review was conducted in three cycles over an 18-month period (September 2004 to March 2006). Search strategies for the literature review initially were designed to cast a wide net but later were refined by using publication type limits and additional terms to produce results that more closely matched the framework of topics and subtopics selected by the Expert Panel. The searches included human studies with abstracts that were published in English in peer reviewed medical journals in the MEDLINE database. Two timeframes were used for the searches, dependent on topic: January 1, 2001, through March 15, 2006, for pharmacotherapy (medications), peak flow monitoring, and written action plans, because these topics were recently reviewed in the EPR-Update 2002; and January 1, 1997, through March 15, 2006, for all other topics, because these topics were last reviewed in the EPR-2 1997.

Search Strategies

Panel members identified, with input from a librarian, key text words for each of the four components of care. A separate search strategy was developed for each of the four components and various key subtopics when deemed appropriate. The key text words and Medical Subject Headings (MeSH) terms that were used to develop each search string are found in an appendix posted on the NHLBI Web site.

Literature Review Process

The systematic review covered a wide range of topics. Although the overarching framework for the review was based on the four essential components of asthma care, multiple subtopics were associated with each component. To organize a review of such an expanse, the Panel was divided into 10 committees, with about 4-7 reviewers in each (all reviewers were assigned to 2 or more committees). Within each committee, teams of two ("topic teams") were assigned as leads to cover specific topics. A system of independent review and vote by each of the two team reviewers was used at each step of the literature review process to identify studies to include in the guidelines update. The initial step in the literature review process was to screen titles from the searches for relevancy in updating content of the guidelines, followed by reviews of abstracts of the relevant titles to identify those studies meriting full-text review based on relevance to the guidelines and study quality.

The combined number of titles screened from cycles 1, 2, and 3 was 15,444. The number of abstracts and articles reviewed for all three cycles was 4,747. Of these, 2,863 were voted to the abstract Keep list following the abstract-review step. A database of these abstracts is posted on the NHLBI Web site. Of these abstracts, 2,122 were advanced for full-text review, which resulted in 1,654 articles serving as a bibliography of references used to update the guidelines, available on the NHLBI Web site. Articles were selected from this bibliography for evidence tables and/or citation in the text. In addition, articles reporting new and particularly relevant findings and published after March 2006 were identified by Panel members during the writing period (March 2006-December 2006) and by comments received from the public review in February 2007.

Preparation Of Evidence Tables

Evidence tables were prepared for selected topics. It was not feasible to generate evidence tables for every topic in the guidelines. Furthermore, many topics did not have a sufficient body of evidence or a sufficient number of high-quality studies to warrant the preparation of a table.

The Panel decided to prepare evidence tables on those topics for which an evidence table would be particularly useful to assess the weight of the evidence-e.g., topics with numerous articles, conflicting evidence, or which addressed questions raised frequently by clinicians. Summary findings on topics without evidence tables, however, also are included in the updated guidelines text.

Evidence tables were prepared with the assistance of a methodologist who served as a consultant to the Expert Panel. Within their respective committees, Expert Panel members selected the topics and articles for evidence tables. The evidence tables included all articles that received a "yes" vote from

both the primary and secondary reviewer during the systematic literature review process. The methodologist abstracted the articles to the tables, using a template developed by the Expert Panel. The Expert Panel subsequently reviewed and approved the final evidence tables. A total of 20 tables, comprising 316 articles are included in the current update (see figure 1-1). Evidence tables are posted on the NHLBI Web site.

Ranking The Evidence

The Expert Panel agreed to specify the level of evidence used to justify the recommendations being made. Panel members only included ranking of evidence for recommendations they made based on the scientific literature in the current evidence review. They did not assign evidence rankings to recommendations pulled through from the EPR-2 1997 on topics that are still important to the diagnosis and management of asthma but for which there was little new published literature. These "pull through" recommendations are designated by EPR-2 1997 in parentheses following the first mention of the recommendation. For recommendations that have been either revised or further substantiated on the basis of the evidence review conducted for the EPR-3: Full Report 2007, the level of evidence is indicated in the text in parentheses following first mention of the recommendation. The system used to describe the level of evidence is as follows (Jadad et al. 2000):

- Evidence Category A: Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.
- Evidence Category B: RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.
- Evidence Category C: Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.
- Evidence Category D: Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

In addition to specifying the level of evidence supporting a recommendation, the Expert Panel agreed to indicate the strength of the recommendation. When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel. When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong. This distinction is an effort to address nuances of using evidence ranking systems. For example, a recommendation for which clinical RCT data are not available (e.g., conducting a medical history for symptoms suggestive of asthma) may still be strongly supported by the Panel. Furthermore, the range of evidence that qualifies a definition of "B" or "C" is wide, and the Expert Panel considered this range and the potential implications of a recommendation as they decided how strongly the recommendation should be presented.

Panel Discussion

The first opportunity for discussion of findings occurred within the "topic teams." Teams then presented a summary of their findings during a conference call to all members of their respective committee. A full discussion ensued on each topic, and the committee arrived at a consensus position. Teams then presented their findings and the committee position to the full Expert Panel at an in-person meeting, thereby engaging all Panel members in critical analysis of the evidence and interpretation of the data.

A series of conference calls for each of the 10 committees as well as four in-person Expert Panel meetings (held in October 2004, April 2005, December 2005, and May 2006) were scheduled to facilitate discussion of findings and to dovetail with the three cycles of literature review that occurred over the 18-month period. Potential conflicts of interest were disclosed at the initial meeting.

Report Preparation

Development of the EPR-3: Full Report 2007 was an iterative process of interpreting the evidence, drafting summary statements, and reviewing comments from the various external reviews before completing the final report. In the summer and fall of 2005, the various topic teams, through conference calls and subsequent electronic mail, began drafting their assigned sections of the report. Members of the respective committees reviewed and revised team drafts, also by using conference calls and electronic mail. During the calls, votes were taken to ensure agreement with final conclusions and recommendations. During the December 2005 meeting, Panel members reviewed and discussed all committee drafts.

During the May 2006 meeting, the Panel conducted a thorough review and discussion of the report and reached consensus on the recommendations. For controversial topics, votes were taken to ensure that each individual's opinion was considered. In July, using conference calls and electronic mail, the Panel completed a draft of the EPR-3: Full Report 2007 for submission in July/August to a panel of expert consultants for their review and comments. In response to their comments, a revised draft of the EPR-3: Full Report 2007 was developed and circulated in November to the NAEPP Guidelines Implementation Panel (GIP) for their comment. This draft was also posted on the NHLBI Web site for public comment in February 2007. The Expert Panel considered 721 comments from 140 reviewers. Edits were made to the documents, as appropriate, before the full EPR-3: Full Report 2007 was finalized and published. The EPR-3: Full Report 2007 will be used to develop clinical practice guidelines and practice-based tools as well as educational materials for patients and the public.

In summary, the NAEPP "Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma-Full Report 2007" represents the NAEPP's ongoing effort to keep recommendations for clinical practice up to date and based upon a systematic review of the best available scientific evidence by a Panel of experts, as well as peer review and critique by the collective expertise of external research/science consultants, the NAEPP CC members, guidelines implementation specialists, and public comment. The relationship between guidelines and clinical research is a dynamic one, and the NAEPP recognizes that the task of keeping guidelines' recommendations up to date is an increasing challenge. In 1991, many recommendations were based on expert opinion because there were only limited randomized clinical trials in adults, and almost none in children, that adequately tested clinical interventions grounded in research findings about the disease process in asthma. The large gaps in the literature defined pressing clinical research questions that have now been vigorously addressed by the scientific community, as the size of the literature reviewed for the current report attests. The NAEPP is grateful to all of the Expert Panel members for meeting the challenge with tremendous dedication and to Dr. William Busse for his outstanding leadership. The NAEPP would particularly like to acknowledge the contributions of Dr. Gail Shapiro, who served on NAEPP Expert Panels from 1991 until her death in August 2006. Dr. Shapiro provided valuable continuity to the Panel's deliberations while simultaneously offering a fresh perspective that was rooted in observations from her clinical practice and was supported and substantiated by her clinical research and indepth understanding of the literature. Dr. Shapiro had a passion for improving asthma care and an unwavering commitment to develop evidence-based recommendations that would also be practical. Dr. Shapiro inspired in others the essence of what NAEPP hopes to offer with this updated Expert Panel Report: a clear vision for clinicians and patients to work together to achieve asthma control.

References

1. EPR. Expert panel report: guidelines for the diagnosis and management of asthma (EPR 1991). NIH Publication No. 91-3642. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, 1991.
2. EPR-2. Expert panel report 2: guidelines for the diagnosis and management of asthma (EPR-2 1997). NIH Publication No. 97-4051. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, 1997.
3. EPR-Update 2002. Expert panel report: guidelines for the diagnosis and management of asthma. Update on selected topics 2002 (EPR-Update 2002). NIH Publication No. 02 5074. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, June 2003.
4. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000;320(7234):537-40.
5. NHIS. National health interview survey (NHIS 2005). Hyattsville, MD: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, 2005. Available at http://www.cdc.gov/nchs/about/major/nhis/reports_2005.htm.

Link to the evidence tables themselves:

http://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/evidence-tables

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

In addition to the work cited above, we conducted a scoping review as follows:

We identified key constructs of asthma ED use measures for consideration. We created a table of these constructs in technical and lay language, and listed research questions for the review to answer. Our contractor (a national accrediting body experienced in measure development), prepared for us a literature review in 2 stages and we supplemented this with targeted reviews as needed to answer specific questions that arose during the measure development process.

The following construct table was used to guide the review and was the basis for the first round of review. Following the table, we include a list of questions for focused review that guided round 2 of the review, which resulted in a detailed summary of 91 articles from the peer-reviewed literature. In addition to this review, the CAPQuaM scientific team conducted an ad hoc series of reviews to answer specific questions such as the reliability of administrative data to identify asthma, and the value of expert panels and the RAND/UCLA appropriateness method. The CAPQuaM degree 360 method starts with a topic area and the measures emerge during the process, in this case necessitating the specified ad hoc reviews.

We searched peer reviewed and gray literature from 1985-2014 over the course of these reviews. Literature was summarized for our expert panel, which met in late 2013.

Overarching statement: Even when not specifically indicated, we are interested in how these constructs are impacted by such factors as race, ethnicity, socioeconomic status or its indicators, or the presence of other special health care needs.

Our metric is designed to capture axes related to two distinct conceptual frameworks:

1) Asthma is a model of chronic disease management. In other words, ED visits may arise from acute exacerbations indicating a flare up of disease, and/or suboptimal management of the chronic illness.
2) ED visits for asthma may reflect limitations of primary care beyond the provision of suboptimal treatment, such as insufficient education, limitations of access or availability, breakdowns of communication, or a variety of other factors.

We note that the internal quality of the ED visit to manage the asthma is not the target of this measure. However, communication between the emergency department and the primary care site may prove to be within the scope of this measure, pending the views of our experts and developers.

Construct I: Need to sufficiently specify population for measure

Concept	Implications (Lay Statement)	Lit Review Questions

(Descriptive) **The measure will need to adequately specify the population that we consider to be eligible for an ED with asthma measure.**	The development of measures regarding ED use for children with asthma requires us to understand the strengths and weaknesses for our measure of various approaches to identifying whether or not children have asthma. It further requires us to understand the impact of the availability of various sources of data (such as encounter data, pharmaceutical data, electronic medical record or chart review data) on these strengths and weaknesses. We are aware that the use of the term asthma is variable. We are not interested in diagnoses with the name asthma, but with an operational diagnosis that we will functionally treat as asthma, whether it has been called chronic wheezing, reactive airway disease, chronic infectious bronchitis, etc. We recognize that asthma and its presentation may change over the course of a child's life.	1. When asthma care is evaluated, how is the population of care recipients defined? How is asthma defined? What is the impact of including various types of data (dx 1 or more, drugs, etc) on the sensitivity and specificity of asthma identification? What are practical and valid approaches to identifying asthma? How do the answers to these questions differ between adults and children? 2. Are any groups persistently excluded from studies of asthma care (i.e., are children who have asthma and other comorbid conditions, such as a malignant disease, excluded?). What rationale is provided for the exclusion? 3. Are any non-asthma diagnoses considered to be indicators of asthma or potential asthma (e.g. bronchitis, bronchiolitis, wheezing, atopy) 4. For children up to age 21, how do issues of diagnosis, management, and follow-up differ by age and developmental stage? 5. At what point does literature suggest that reactive airway disease should be managed as asthma? a. What other conditions are managed as asthma? 6. What common current or preexisting comorbid conditions alter the management plan for asthma?

Construct II: Adequacy of management of asthma (as a chronic disease example)

Concept	Implications (Lay Statement)	Lit Review Questions
IIA. **↑Adequacy of asthma management:** **↓ED visits**	Since asthma is a chronic disease characterized by acute exacerbations, the extent to which asthma care is optimized through the use of appropriate medications, the control of the environment, and	1. What are the recommendations of the NHLBI guidelines? a. What does the literature suggest about the usefulness of NHLBI guidelines? b. Are there aspects that it has identified that appear to be missed? 2. What do we know about asthma management, how it's measured,

	the preparation of the parent/child dyad to adapt to changes in circumstances (e.g. viral respiratory infection or exposure to cold) should reduce the number of ED visits, irrespective of the number of primary care visits.	who provides it, patterns of care and how ED visits vary as a consequence? 3. Does identification of PCP improve outcomes of ED visit, including patterns of care, utilization? 4. What do we know about the content of an asthma plan and its relationship to a full program of chronic disease management, and its influence on ED utilization? 5. What evidence is there about the impact on outcomes such as ED use when the child or adolescent is involved in asthma self-management? For example, does it matter if: a. The child has a written asthma plan? b. The child understands their asthma plan? c. The child is given an opportunity to participate in managing care? 6. How is the role of the child in self-management measured? 7. How much are children able to recognize, communicate and act on their asthma? 8. What do we know about the impact of asthma services on asthma management? This includes: a. Treatment from an asthma specialist; b. Social worker; or c. Multidisciplinary personnel 9. To what extent is ED use by children with asthma stimulated by non-asthma related issues? a. How can we identify when that occurs? b. What is the evidence that providing other services will reduce the number of ED visits? 10. To what extent do children contribute to their management (including avoiding triggers, recognizing symptoms, medication adherence, etc.)? a. What is the impact and variance by age?

	11. What is the evidence regarding adequacy of various medication delivery systems for infants, toddlers, children and adolescents in acute and chronic settings? 12. Is there evidence of prior insult to the lungs such as sequelae of prematurity, etc. that create distinct subpopulations when considering this measure (at risk for ER visit)? 13. What aspects of the health services environment have been identified as contributing to outcomes of asthma management (e.g. school based health care)? 14. Does rate of ED utilization for non-respiratory diagnoses vary between asthmatics and non-asthmatics? 15. What is known about how often children with asthma use the ED over an extended period of time? Does it change over the life course of childhood? How does that vary by child characteristics, including race, SES, urban, suburban vs. rural, and age?

IIB. **↑PCP capacity/knowledge/skill:** **a. ↑Asthma management** **b. ↓Asthma exacerbations** **c. ↑Chronic disease management**	Broadly speaking, patient management of asthma is influenced by the capacity of the PCP practice. This includes the knowledge and skills possessed by the PCP, as well as office support to enhance access and availability of care. PCP includes the ability of the PC office to meet the cultural needs of the patient and their family.	1. What are the diversity of practices or services that may or may not impact ability or capacity of the PCP practice to manage asthma? 2. What do we know about the specific skills and processes that contribute to a primary care practice's capacity? 3. What patterns of visits or medication use or other indicators have been used as markers of well or poorly delivered primary care for asthma in children and/or adults? 4. What is the minimum use of specialists appropriate for children with asthma? How does that vary with history of ED or hospital use? a. When and how does the use of specialists become a marker for higher or lower quality of care? 5. What evidence is there regarding the nature of the PCP practice for children with asthma? For example, the level of continuity with individual clinicians vs. practices, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc.

IIC. **↑Asthma education:** **a. increases recognition of symptoms >** **b. ↑Management skills**	Enhancing what patients or their families know about asthma may be an important tool to improve care for children with asthma. The likely first effect of such education is to enhance the capacity of a caregiver to identify what symptoms may relate to asthma. This could conceivably increase utilization of both PCP and ED services if this were to increase the caregiver's perceived need for care for their child's asthma. With a more sophisticated understanding, including having a valid asthma action plan and understanding how to use it, ED care may be reduced and PCP care for asthma may be reduced, as symptoms are less frequent and parents are more competent to manage them when they arise.	1. What are metrics or processes regarding the quality of asthma care? Is it drug ratios (i.e. proportion of prescriptions filled that are for rescue vs control medications), asthma action plan, , capacity of PCP office, relationship to PCP practice, or other specific bundles of care, etc? 2. What constitutes "perfect care"/"best practice" for any specified type of patient? 3. What do we know about the impact of asthma education programs on quality of care, outcomes of care, or utilization of care? Define utilization of care as including: a. PCP utilization, b. ED utilization, c. Referral/specialist utilization, d. Non physician care team member utilization, e. Medication usage, f. Hospitalizations, and/or g. Other care utilization areas to consider? Examples may include functional status, quality of life elements, spirometry, role functioning. 4. What is the diversity of asthma education programs and what are the differences in quality of care/outcomes/utilization of care associated with differences? 5. Does referral to an asthma specialist impact quality of care, utilization of care and asthma outcomes? 6. Does referral to a social worker impact utilization of care and asthma outcomes? 7. (Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes? 8. What are desirable roles and effectiveness of interventions that extend beyond the healthcare system, such as reducing pollution, focusing on environmental justice, housing, dust mites, etc.? 9. How does organization and capacity of the practice setting influence the delivery of asthma management education?

Construct III:	Adequacy of PCP practice site to handle acute exacerbations of chronic disease and/or acute illnesses	
Concept	**Implications (Lay Statement)**	**Lit Review Questions**
IIIA. **↑Primary care capacity:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)** **IIIA.2** **SUBCONSTRUCT:** **↑Accessibility:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	In general, enhanced capacity may affect a patient's access to care. Capacity can refer to patient services that make it easier for a patient to receive timely care, such as location or hours of offices, to the ability to triage phone calls in a timely and effective way, or may include the materials and services present within an office (e.g. the presence of a treatment room, the capacity to deliver oxygen, nebulizers, etc.) Such capacity may be limited or enhanced by staffing, space, the ability to safely transport someone from the office to a hospital, etc. If PCP office capacity is optimized, ED visits may be reduced as acute and mundane conditions can be managed in a PCP setting. Subsequently, increased capacity of the entire PCP support network will increase number of PCP visits.	1. What do we know about access to the PCP's office as a place to manage asthma, and the subsequent capacity of a PCP and the diversity of practice settings? Additionally, how do we measure capacity and, its impact on QoC, processes of care, asthma outcomes, asthma specific processes and utilization? How do these factors impact ED use or other outcomes? a. In general: i. PCP/specialist ratio in a plan or PCP/child ratio ii. PCP time spent in visit (incl. minutes per sick, well-child, asthma management visit) iii. Nature of training activities iv. How long does it take to schedule a visit (incl. asthma (chronic), acute, follow-up visit) v. Office hours and visit flexibility (incl. after hours coverage, office consult, meet in ED) vi. Phone capabilities: (incl. answering capacity, putting on hold, returning calls, after hours phone service) vii. Level of implementation of patient centered medical home/chronic care model, eg i. Use of registries ii. Standardized tools for measurement iii. Case management iv. Group visits or other education, etc b. Specifically, ability to manage acute dx in office, which includes: i. Do they have a treatment room or capacity to use a room as a treatment room? ii. Do they offer rescue treatments (e.g. nebulizers, spacers)? iii. Can they measure oxygen saturation?

iv. Do doctors feel comfortable with acute asthmatic in office?
v. Can they take time to manage an acute pt in their office?
vi. Do they have safe and rapid transport to a hospital (how long?)

2. Availability and accessibility of offices (incl. office hours, geographic distribution)
   a. What do we know about linguistic capabilities in the PCP setting influencing use of the ED?
   b. What do we know about proximity of the PCP office to public transit on the utilization of the ED?
3. What do we know about the impact of variations in patterns of care/practice, use of modalities, and/or and receipt of well-child care on asthma management or outcomes (eg ED use)? Does Immunization status reflect on t eh capacity of the PCP, on the state of the child, or on other factors that may relate to asthma outcomes? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?
4. Are children with more WCC visits less likely to use the ED for acute visits? children who are UTD on their immunizations?
5. What literature is there on the relationship between pediatric ED use and other measures of asthma exacerbation/outcomes?
6. What do we know about variability of capacity and management of mundane conditions (e.g. OM, URIs, pharyngitis), office to ED ratios?
7. What do we know about variability of capacity and management of acute conditions requiring interventions (e.g. asthma)?
8. To what extent does ED capacity increase use of ED services? Do hospitals advertise ED services, have fast track for mundane conditions, etc?
9. To what extent does ED have capacity to provide primary care, routine immunizations, etc? How is that built into policies and

	protocols? 10. At what age does the PCP start meeting alone with child? Time spent in visit? 11. To what extent and at what age do PCP's involve children in self-management and does it vary?

IIIB. **↑Relationship with PCP:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	Improved relationship with PCP may increase visits to your PCP and decrease ED visits, for both acute and mundane conditions. A good relationship may lead to greater trust and adherence to recommendations (both WCC and asthma care) and drive a preference for seeking care by the PCP over seeking care in another environment. In general, we are referring to relationship of caregiver with PCP and their office staff. We recognize the importance of the relationship of PCP's with patients as well; when the relationship between the PCP and the child rather than caretaker is emphasized in research, we'd like to capture that as well.	1. What exists regarding measuring the quantity and quality of the relationship with PCP? Specifically: a. What's the variation and does it matter? b. How is it measured? c. What do we know about patient experience of care, especially as it relates to relationship with clinicians/PCP d. To what extent is quality of relationship expressed in terms of caregiver vs. child relationships and how does this change with age of child or longevity of connection to a PCP? 2. What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care? Define supplemental services as: a. Electronic educational/reminder tools (incl. social media) b. Telephone educational/reminder tools c. Print materials (e.g. educational brochures) d. Disease management, demand management, or other type programs e. Other services to consider? Measure quality, utilization of care should include at least : a. ED visits b. PCP visits 3. How does role of child in self care/management tie into these issues?

Construct IV: The connectedness of care in the primary care and ED setting – before, during, and after of the ED visit

Concept	Implications (Lay Statement)	Lit Review Questions
IV. (Descriptive) **Enhanced integration of ED care of asthma with routine care will have better outcomes**	If primary care is generally pretty good, then the ED visit should be an extraordinary event. In such cases the PCP alerting the ED to current management and the ED assuring appropriate follow up with the PCP is important. In cases where primary care is of lower quality or more variable, the ED visit may enhance the long term management of the child with asthma. And we need to assess this. One of the ways it might do so is to construct an asthma management plan that is then followed by the PCP. Another way is to connect a child without adequate primary care to primary care, especially to someone who is competent to manage the asthma.	1. What evidence supports that ED visits for asthma are most effective when visit is followed by a visit to the PCP? 2. Do utilization patterns in both the ED and primary care setting change following ED visits? 3. Is an effective/more effective use of medications seen following an ED visit? 4. Does the identification of a primary care provider improve outcomes of an ED visit (including patterns of care utilization)? 5. Is pre or intra visit communication with the primary care provider associated with better outcomes? How often does this occur? Are there systematic differences regarding those for whom this does and does not occur? 6. Are ED visits for asthma routinely associated with some form of communication or linkage with PCP? Does that result in better outcomes?

Construct V: Equity is a value in asthma care

Concept	Implications (Lay Statement)	Lit Review Questions
V. (Descriptive) **Equity is a critical construct of quality for children with equity**	Systematic differences in the frequency or nature of ED visits for asthma on the basis of race, ethnicity, family make-up, income/economic status, specifics of insurance status, presence or absence of comorbid special health care needs, etc represents decrements in quality that our measures should identify.	1. Does the literature indicate systematic or predictable differences in the frequency or nature of asthma care for children as it relates to ED visits for asthma that may be interpreted as representing inequitable structures, processes, outcomes, experiences with, or coordination of care? 2. What do we know about how social determinants and diagnosis and management of asthma and its outcomes, specifically as it relates to use of ED? 3. What do we know about the extent to which use of the ED for children with asthma that relates to the external physical and social environment?

Proposed Research Questions

Asthma- We propose to prioritize our Asthma Construct Table, to the following questions:

Acronyms

PCP: Primary Care Provider

ED: Emergency Department

WCC: Well-child care

Baseline Question (for Questions 1, 2 and 3 below):
When asthma care is evaluated, how is the population of asthma care recip
population level? What are specific implications of how you identify patie
including various approaches to specifying the denominator of children wi
practical and valid approaches to identifying asthma at the population level? How do the answers to these questions differ between adults and children?

Question 1 (Construct IIA.2):

For children with asthma, what do we know about asthma management? How is management of asthma described and measured? This includes who (PCP, asthma specialist, ED, etc) primarily manages it as well as who provides it. What are the patterns of care and what do we know about how use of the ED varies as a result of various approaches to management?

- **Question 1a (Construct IIB.3):**
  Specifically, have any of these patterns of visits or medication use or other characteristics of care been used as markers of well or poorly delivered primary care for asthma for children and/or adults?

Question 2 (Construct IIB.5):

How has varying asthma care for children been described on the basis of characteristics of the PCP offices or practices? For example, are they characterized by the level of continuity between individual clinicians, the level of conntiuity with any provider in the practice, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc?

- **Question 2a (Construct IIIA.3):**
  What do we know about the impact of variations in patterns of care/practice, use of treatment modalities, and/or receipt of well-child care on asthma management or outcomes (e.g. ED use)? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?

Question 3 (Construct IIC.7):

(Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes both within context of a primary care practice or in other clinical settings?

- **Question 3a. (Construct IIIB.2):**

What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care?

1a.8.2. Provide the citation and summary for each piece of evidence.

Our approach to developing this measure stems from a vibrant and scientifically sound tradition regarding measuring performance. We discuss herein research involving the soundness administrative data to identify children with asthma. This is a generally accepted and standard approach with acceptable reliability.

Brook and Davies [1] trace the early history of quality measurement and remind us of the importance of medical chart audit as an approach to quality measurement. Lohr and Brook at RAND and Roos in Manitoba, Canada pioneered the use of electronically-available administrative data (generated by routine health care operations, such as billings) as proxies for health care processes. Administrative data carefully used reduces burden of quality measurement. [2-6]

As the National Committee for Quality Assurance (NCQA) developed the Healthcare Employee Data Information Set (HEDIS) as the de facto measurement system for managed care, attention turned to the use of administrative data for routine performance measurement.

We have used rigorous and transparent methods [14] to assemble a national expert panel that included pediatricians, family physicians, pediatric and general emergency room specialists, a pediatric pulmonologist and a pediatric allergist from practices and medical schools around the country. This work was conducted in collaboration with national clinical societies (AAP, AAFP) and CAPQuaM's diverse other partner organizations, including NY State DoH/Medicaid. NCQA is an important technical consultant and partner. The specific criteria that we operationalize in this measure were all rated by the expert panel with a median score of 8 or 9 on a 9 point scale (9 high) to develop inclusion and exclusion criteria, variables for stratification and so forth. The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children.

The literature has demonstrated the reliability of claims data for assessing asthma. Though they have their limitations, these data types have been shown in multiple studies to be a reliable source of information for population level quality measurement. They are currently used for all of the analogous measures of which we are aware, including the former Core Measure and the NCQA measure considering children with persistent asthma.

The use of two years of data to validate the diagnosis of asthma has been found to produce substantial agreement with patient surveys and improves performance over the use of one year of data (28). Others have reported that using administrative databases to identify asthma is both sensitive and specific as compared to review of the primary care physician's office chart (29).

Select additional references documenting other aspects of performance gap, and supporting our process and data sources are also noted (7-13, 15-35).

1. Brook, R.H. and A. Davies-Avery, *Mechanisms for Assuring Quality of US Medical Care Services: Past, Present, and Future*. 1977: Santa Monica, CA.

2. Brook, R.H. and K.N. Lohr, *Efficacy, effectiveness, variations, and quality. Boundary-crossing research.* Med Care, 1985. **23**(5): p. 710-22.

3. Lohr, K.N., *Use of insurance claims data in measuring quality of care.* Int J Technol Assess Health Care, 1990. **6**(2): p. 263-71.

4. Roos, L.L., et al., *Using Administrative Data Banks for Research and Evaluation: A Case Study.* Evaluative Research 1979. **236**(3).

5. Roos, L.L., Jr., J.P. Nicol, and S.M. Cageorge, *Using administrative data for longitudinal research: comparisons with primary data collection.* J Chronic Dis,1987. **40**(1): p. 41-9.

6. Roos, L.L., Jr., et al., *How good are the data? Reliability of one health care data bank.* Med Care, 1982. **20**(3): p. 266-76.

7. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.

8. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p.

199-207.

9. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p.

813-25.

10. Angier, H., et al., *Variation in outcomes of quality measurement by data source.* Pediatrics, 2014. **133**(6): p. e1676-82.

11. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.

12. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.

13. NCQA. *National Committee for Quality Assurance.* [cited 2014 7/30/14];

Available from: http://www.ncqa.org/.

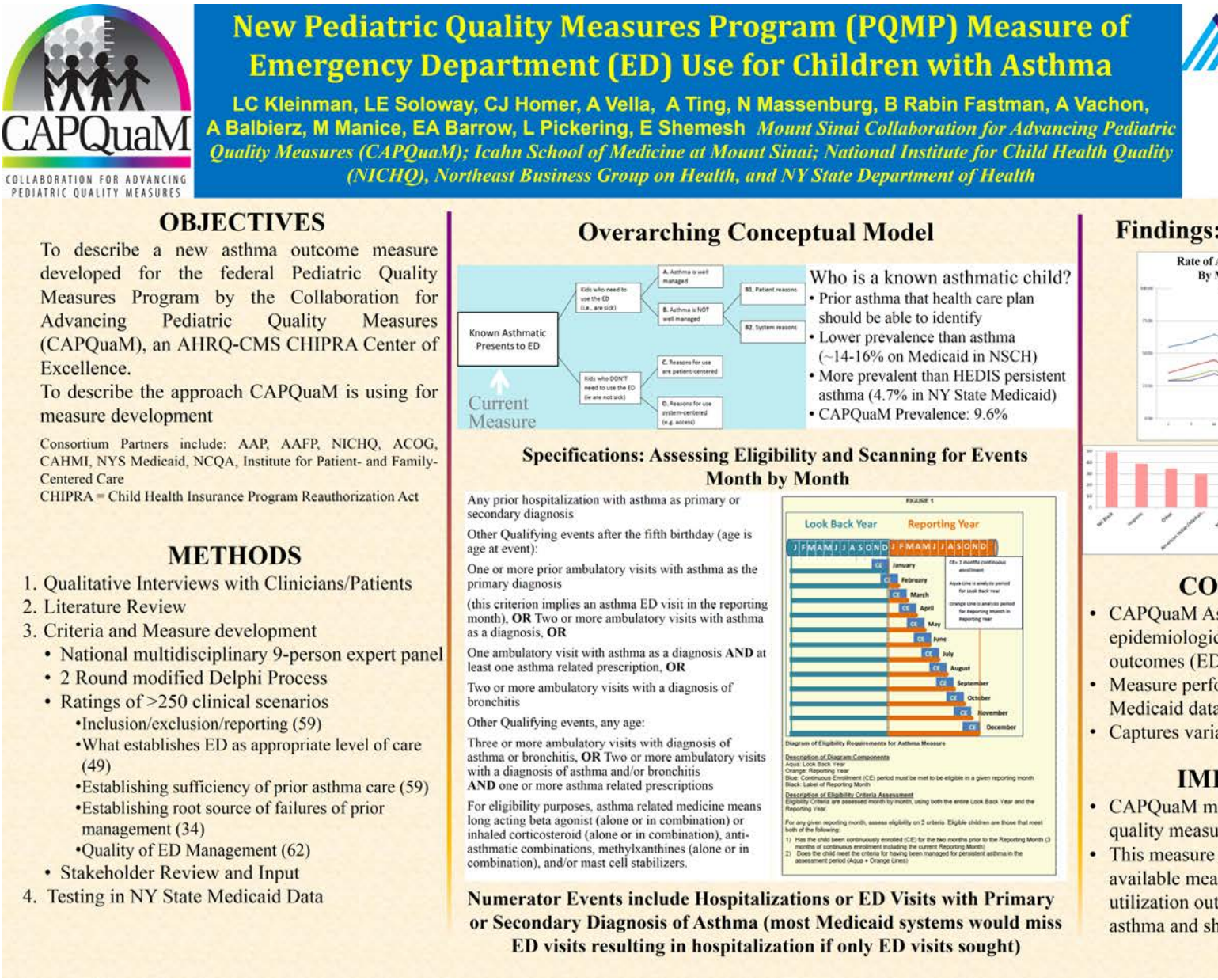
14. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.

15. DeAngelis, C., P. Fosarelli, and A.K. Duggan, *Use of the emergency department by children enrolled in a primary care clinic.* Pediatr Emerg Care, 1985. **1**(2): p.61-5.

16. Talreja, N., et al., *Modifiable factors associated with severe asthma exacerbations in urban patients.* Ann Allergy Asthma Immunol, 2012. **109**(2): p.128-32.

17. Auger, K.A., et al., *Medical home quality and readmission risk for children hospitalized with asthma exacerbations.* Pediatrics, 2013. **131**(1): p. 64-70.
18. Ducharme, F.M., et al., *Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control.* Am J Respir Crit Care Med, 2011. **183**(2): p. 195-203.
19. Farber, H.J., *Optimizing maintenance therapy in pediatric asthma.* Curr Opin Pulm Med, 2010. **16**(1): p. 25-30.
20. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, *Relationship between pediatric primary provider visits and acute asthma ED visits.* Pediatr Pulmonol, 2007. **42**(11): p. 1041-7.
21. Parchman, M.L. and S. Culler, *Primary care physicians and avoidable hospitalizations.* J Fam Pract, 1994. **39**(2): p. 123-8.
22. Prevention, C.f.D.C.a. *Home-based Multi-trigger, Multi-component interventions*. 2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
23. Price, J.H., et al., *Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States.* Biomed Res Int, 2013. **2013**: p. 1-12; Available in open access http://dx.doi.org/10.1155/2013/787616
24. Homer, C.J., et al., *Does quality of care affect rates of hospitalization for childhood asthma?* Pediatrics, 1996. **98**(1): p. 18-23.
25. Finkelstein, J.A., et al., *Quality of care for preschool children with asthma: the role of social factors and practice setting.* Pediatrics, 1995. **95**(3): p. 389-94.
26. Flores, G., et al., *Keeping children out of hospitals: parents' and physicians' perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided.* Pediatrics, 2003. **112**(5): p. 1021-30.
27. Knudson, A., et al., *Disparities in pediatric asthma hospitalizations.* J Public Health Manag Pract, 2009. **15**(3): p. 232-7.
28. Bindman, A.B., et al., *Preventable hospitalizations and access to health care.* Jama, 1995. **274**(4): p. 305-11.
29. National Heart, Lung and Blood Institute. *Asthma Guidelines*. 2011 February 2011 [cited 2014 7/30/2014].
30. *A Matter of Urgency: Reducing Emergency Department Overuse* in *A NEHI Research Brief* 2010, New England Healthcare Institute.
31. Martin, B.C., *Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital.* J Health Care Finance, 2000. **27**(2): p. 51-65.
32. Finkelstein, J.A., et al., *Comparing asthma care for Medicaid and non-Medicaid children in a health maintenance organization.* Arch Pediatr Adolesc Med, 2000. **154**(6): p. 563-8.
33. Owens, P.L., et al., *Care of children and adolescents in U.S. hospitals*, in *HCUP Fact Book No. 4*. Agency for Healthcare Research and Quality: Rockville: MD.
34. Pearson, W.S., et al., *State-based Medicaid costs for pediatric asthma emergency department visits.* Prev Chronic Dis, 2014. **11**: p. E108.
35. Taubman, S.L., et al., *Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment.* Science, 2014. **343**(6168): p. 263-268.

The following poster describing this measure was submitted for peer review and accepted and presented at the Annual Research Meeting of AcademyHealth in 2014.



1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

evidence_attachment_asthma_ED_Rate.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

ED visits for children with asthma is an outcome measure of intrinsic value. It represents utilization of an expensive service and constitutes a burden on children and their families. There is abundant evidence that ED visits are common, may be reduced through improved primary care or community-based interventions, and demonstrate disparities (1-11, 12-19). Asthma is generally recognized to be an ambulatory care sensitive condition. Nonetheless, we perceive and our panel articulated that the rate for ED visits ought not to be 0. So while in general a lower rate represents preferable care, too low a rate could indicate insufficient access to emergency room services. Our overarching conceptual framework that extends beyond this measure is shown in the evidence form.

Our measure benefits from a formal development process, CAPQuaM's 360 degree method, which is described in more detail in the measure testing form. The measure and its specifications result from a formal development process for this measure incorporated stakeholder input including a parent focus group, meeting with The Mount Sinai Pediatrics Department's Parent Advisory Council, interviews with primary care clinicians and ED physicians, the CAPQuaM's multidisciplinary scientific team, which includes investigators, a steering committee and a senior advisory board of nationally prominent figures. The measure also benefits from a national multidisciplinary Expert Panel which utilized a RAND type modified Delphi method to guide our specifications.

When epidemiologists describe how frequently something occurs the preferred measure is typically an incidence density, or rate. In contrast to a risk or proportion, the incidence density has as its denominator a measure of the extent of potential exposure in the population, expressed in people-years. This measure represents an advance in the measurement of healthcare performance for children: it incorporates this formulation both to enhance its interpretation (because it has a specific epidemiological meaning) and to limit distortion if sick children move in or out of eligibility for the measure. (20)
Further clinical evidence of gaps are demonstrated in the description by NHLBI's NAELPP guideline, cited in the evidence form, Schatz and colleagues study describing the relationship between asthma control and asthma exacerbations in managed care (21), and Fuhlbrigge et al's confirmation that medications can work to reduce ED visits for asthma but are used sub optimally (22).

References

1. Adams, R.J., B.J. Smith, and R.E. Ruffin, Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. Thorax, 2000. 55(7): p. 566-73.
2. Bahadori, K., et al., Economic burden of asthma: a systematic review. BMC Pulm Med, 2009. 9: p. 24.
3. Weiss, K.B., P.J. Gergen, and T.A. Hodgson, An economic evaluation of asthma in the United States. N Engl J Med, 1992. 326(13): p. 862-6.
4. Coventry, J.A., M.S. Weston, and P.M. Collins, Emergency room encounters of pediatric patients with asthma: cost comparisons with other treatment settings. JAmbul Care Manage, 1996. 19(2): p. 9-21.
5. Association, A.L. American Lung Association Fact Sheet on Asthma and Children. October 2012 [cited 2013 May 20]; Available from: http://www.lung.org/lung-disease/asthma/resources/facts-and-figures/asthmachildren-fact-sheet.html#5.
6. Fuhrman, C., et al., Hospitalizations for asthma in children are linked to undertreatment and insufficient asthma education. J Asthma, 2011. 48(6): p. 565-71.
7. Sawicki, G.S., et al., Uncontrolled asthma in a commercially insured population from 2002 to 2007: trends, predictors, and costs. J Asthma, 2010. 47(5): p. 574-80.
8. Manice, M., Exploring the relationship between parental shared decision-making practices and acute asthma exacerbations among children age 0-17. 2013, Icahn School of Medicine at Mount Sinai: New York, NY.
9. Cerdan, N.S., et al., Asthma severity in children and the quality of life of their parents. Appl Nurs Res, 2012. 25(3): p. 131-7.
10. Fiese, B., et al., Family climate of routine asthma care: associating perceived burden and mother-child interaction patterns to child well-being. Fam Process, 2008. 47(1): p. 63-79.
11. Okelo, S.O., et al., Emotional quality-of-life and outcomes in adolescents with asthma. J Pediatr, 2004. 145(4): p. 523-9.
12. Talreja, N., et al., Modifiable factors associated with severe asthma exacerbations in urban patients. Ann Allergy Asthma Immunol, 2012. 109(2): p.128-32.
13. Auger, K.A., et al., Medical home quality and readmission risk for children hospitalized with asthma exacerbations. Pediatrics, 2013. 131(1): p. 64-70.
14. Ducharme, F.M., et al., Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control. Am J Respir Crit Care Med, 2011.183(2): p. 195-203.
15. Farber, H.J., Optimizing maintenance therapy in pediatric asthma. Curr Opin Pulm Med, 2010. 16(1): p. 25-30.
16. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, Relationship between pediatric primary provider visits and acute asthma ED visits. Pediatr Pulmonol, 2007.42(11): p. 1041-7.
17. Self, T.H., et al., Reducing emergency department visits and hospitalizations in African American and Hispanic patients with asthma: a 15-year review. J Asthma, 2005. 42(10): p. 807-12.

18. Lara, M., et al., Reducing quality-of-care disparities in childhood asthma: La Red de Asma Infantil intervention in San Juan, Puerto Rico. Pediatrics, 2013. 131 Suppl 1: p. S26-37.
19. Oraka, E., et al., Racial and ethnic disparities in current asthma and emergency department visits: findings from the national health interview survey, 2001-2010. J Asthma, 2013. 50(5): p. 488-96.
20. Rothman, K.J., Modern Epidemiology. 3rd edition ed. 2008, Philadelphia, PA: Lippincott, Williams and Wilkins.
21. Schatz M, Zeiger RS, Yang ST, et al. "Relationship of asthma control to asthma exacerbations using surrogate markers within a managed care database." Am J Managed Care. 16(5):327-333, 2010.
22. Fuhlbrigge A, Carey VJ, Adams RJ. "Evaluation of asthma prescription measures and health system performance based on emergency department utilization." Med Care 42(5):1-7, 2004.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The overall rate of ED visits for asthma in NY State Medicaid Managed Care in 2012 is 28.95 per 100 child-years in children 2-18 and 29.25 for children 2-21. By age stratum the rates are 47.4 visits per 100 child-years for children 2-4, 26.0 for children 5-11, 22.7 for adolescents 12-18, and 34.1 for adolescents 19-21. The appendix contains additional data including demonstrating expected seasonal variations in rate. Given our findings and our methods, although we consider this measure to be specified for a year we have demonstrated its validity to identify or compare asthma ED rate on a month-by-month basis.

The Appendix includes more data as indicated:

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Asthma is a critical problem with racial and ethnic disparities and varies by urbanicity. Adherence to the National Asthma Education and Prevention Programs (NAEPP) Guidelines improves outcomes. [1-32]. We have elsewhere provided other articles, studies, and summaries of evidence to document that ED visits and hospitalizations are typically outcome measures of choice when assessing asthma control.

1. PCORI. PCORI Funding Annoouncement: Treatment Options for Afircan Americans and Hispanics/Latinos with Uncontrolled Asthma. 2013 [cited 2013 September 18]; Available from: http://pcori.org/assets/2013/06/PCORI-Asthma-PFA-061813.pdf.

2. Marcano-Belisario, J., Greenfield G, Huckvale K, Gunn LH, Car J, Apps for asthma self-management: a systematic assessment of content and tools. Cochrane Database Syst. Rev., 2012(8).

3. Health, O.o.M. Asthma and African Americans. [Fact Sheet]. 2012 [cited 2013 August 28]; Available from: http://minorityhealth.hhs.gov/templates/content.aspx?ID=6170.

4. Health, O.o.M. Asthma and Hispanic Americans. [Fact Sheet]. 2012 [cited 2013 August 28]; Available from: http://minorityhealth.hhs.gov/templates/content.aspx?ID=6173.

5. Wennergren, G., Strannegard I, Asthma hospitalizations continue to decrease in schoolchildren but hospitalization rates for wheezing illnesses remain high in young children. Acta Paediatr, 2002. 91(11): p. 1239-1245.

6. Wisnivesky, J., Lorenzo J, Lyn-Cook R, et al., Barriers to adherence to asthma management guidelines among inner-city primary care providers. Annals of Allergy, Asthma & Immunology, 2008. 101(3): p. 264-270.

7. DiSantostefano, R., Davis K, Yancey S, Crim C, Ecologic analysis of asthma-related events and dispensing of inhaled corticosteroid- and salmeterol-containing products. Ann Allergy Asthma Immunol, 2008. 100(6): p. 558-565.

8. Crocker, D., Kinyota S, et al., Effectiveness of home-based, multi-trigger, multicomponent interventions with an environmental focus for reducing asthma morbidity: a community guide systematic review. Am J Prev Med, 2011. 41(2): p. S5-S32.

9. Gustafson, D., Wise M, et al., The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: a randomized controlled trial. J Med Internet Res., 2012. 14(4): p. e101.

10. Program, T.N.A.E.a.P., Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. J Allergy Clin Immunol, 2007. 120(5 Suppl): p. S94-138.

11. Celano, M., Holsey CN, Kobrynski LJ, Home-based family intervention for low-income children with asthma: a randomized controlled pilot study. J Fam Psychol., 2012. 26(2): p. 171-178.

12. Zuniga, G., Kirk S, et al., The impact of asthma health education for parents of children attending head start centers. J Community Health, 2012. 37(6): p. 1296-1300.

13. Blaakman, S., Tremblay PJ, Halterman JS, Fagnano M, Borrelli B, Implementation of a community-based secondhand smoke reduction intervention for caregivers of urban children with asthma: process evaluation, successes and challenges. Health Educ Res, 2013. 28(1): p. 141-152.

14. Martin, M., Catrambone CD, et al., Improving asthma self-efficacy: developing and testing a pilot community-based asthma intervention for African American adults. J Allergy Clin Immunol, 2009. 123(1): p. 153-159.

15. Seid, M., D´Amico E, Varni JW, et al., The In Vivo Adherence Intervention For at Risk Adolescents With Asthma: Report of a Randomized Pilot Trial. Journal of Pediatric Psychology, 2012. 37(4): p. 390-403.

16. Edwards, J., INSPIRE curriculum delivered in a faith-based setting. Fam Community Health., 2010. 33(2): p. 117-122.PRINCIPAL INVESTIGATOR

17. Press, V., Pappalardo AA, et al., Interventions to improve outcomes for minority adults with asthma: a systematic review. J Gen Intern Med, 2012. 27(8): p. 1001-1015.

18. DeJongh, T., Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R, Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst. Rev., 2012.

19. James, T., Fine M, Monitoring Asthma Control Using Claims Data And Patient-Reported Outcomes Measures. P.T., 2008. 33(8): p. 454-466.

20. Bender, B., Overcoming barriers to nonadherence in asthma treatment. J Allergy Clin Immunol, 2002. 109(6): p. 554-559.

21. Okelo, S., Eakin M, et al., The Pediatric Asthma Control and Communication Instrument asthma questionnaire: for use in diverse children of all ages. J Allergy Clin Immunol, 2013. 132(1): p. 55-62.

22. Benavides, S., Rodriquez JC, Maniscalco-Feichtl M, Pharmacist involvement in improving asthma outcomes in various healthcare settings: 1997 to present. Ann Pharmacother. , 2009. 43(1): p. 85-97.

23. Drotar, D., Physician behavior in the care of pediatric chronic illness: association with health outcomes and treatment adherence. J Dev Behav Pediatr., 2009. 30(33): p. 246-254.

24. Weinstein, A., The potential of asthma adherence management to enhance asthma guidelines. Ann Allergy Asthma Immunol, 2011. 106(4): p. 283-291.

25. Agency, E.P. President's Task Force on Environmental Health Risks and Safety Risks to Children:Coordinated Federal Action Plan to Reduce Racial and Ethnic Asthma Disparities. 2012 [cited 2013 August 28]; Available from: http://www.epa.gov/childrenstaskforce/actionplan_one_pager.pdf.

26. Fisher, E., Strunk RC, et al., A Randomized Controlled Evaluation of the Effect of Community Health Workers on Hospitalization for Asthma. Arch Pediatr Adolesc Med, 2009. 163(3): p. 225-232.

27. Cazzola, M., Matera M, Safety of long-acting beta2-agonists in the treatment of asthma. Ther Adv Respir, 2007. 1(1): p. 35-46.

28. Halterman, J., Fagnano M, et al., The school-based preventive asthma care trial: results of a pilot study. J Pediatr, 2012. 28(1): p. 1109-1115.

29. Association, A.L. Trends in Asthma Mortbidity and Mortality. 2012.

30. Cloutier, M., Hall C, Wakefield D, Bailit H, Use of asthma guidelines by primary care providers to reduce hospitalizations and emergency department visits in poor, minority, urban children. J Pediatr, 2005. 146(5): p. 591-597.

31. Prevention, C.f.D.C.a., Vital Signs: Asthma Prevalence, Disease Characteristics, and Self-Management Education -- United States, 2001--2009. 2011: p. 547-552.

32. Clinical Guidelines for the Diagnosis, Evaluation and Management of Adults and Children with Asthma, in New York State Department of Health and Mental Hygiene. 2008. Lawrence C. Kleinman, MD, MPH, FAAP

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
On a yearly and a monthly basis we can demonstrate differences in the data by age, urbanicity, race/ethnicity, and level of poverty. Such differences are also evident in other cross tabulations, for example, the rate for children 2-4 in large metropolitan areas is 52.6 visits per 100 child-years compared to those in small metropolitan areas with 26.2, in micropolitan areas with 18.3 and in rural areas with 12.3. Similar magnitudes of differences were seen in other age groups, although the patterns were not all identical. Racial and ethnic differences were notable: for children ages 2-4, the rate in non-Hispanic Whites was 18.4 visits per 100 child-years, in Asians 19.3, in Hispanics 53.9 and in non-Hispanic Blacks 74. Although less dramatic, similar patterns were observed in all age groups. Overall, the rate for different races ordered by varying magnitude as illustrated between Black and White children, 41.99 and 14.79, respectively. The rate for Hispanic children was intermediate at 31.91 visits per 100 child-years. Charts and graphs are shown in the Appendix Tables 2-5 and Figures 2-5. Other disparities data has been cited elsewhere in terms of asthma control and outcomes.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
See 1.b.4 and Appendix.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Please see both literature review and data presented above. ED visits for asthma in children are common and expensive. They may result from poor quality of care delivered (failure to adhere to guidelines) as well as from insufficient access to primary care. Asthma is the leading diagnosis leading to urgent care/emergent care provided in emergency departments for children. It is among the most common chronic diseases in children and expenses for asthma care are in the billions of dollars annually. Further, CMS and AHRQ assigned us this measure. In addition to data and citations provided, the team has analyzed 2007 and 2011 waves of the National Survey of Children's Health and confirmed that this parent reported measure both identified a high prevalence of asthma nationwide and significant consequences in terms of parent reported child health for children who have asthma.

Our analysis of National Survey of Children's Health [33] data (NSCH, 2011/12), estimates that 10.3 million children in the U.S. have been told that they have asthma. Of these children 7.6 million live in more urban areas that are characterized as metropolitan statistical areas (MSAs), an asthma prevalence rate of 15.4%. Table 1 shows that asthma is very consequential for health.

Table 1. Impacts of Asthma for Children Age 2-17, NSCH 2011/12
Parent/caregiver reports child's health status is excellent or very good

	2 - 5 years	6 – 11 years	12 – 17 years	Total
All Children living in Metropolitan Statistical Areas				
Asthma	59.8 %	69.6 %	74.3 %	70.1 %
No asthma	87.8 %	85.3 %	85.1 %	85.9 %
Overall	84.9 %	82.8 %	83.1 %	83.4 %
Difference	-28.0 %	-15.7 %	-10.8 %	-15.8 %
Children living in MSAs with Asthma				
All Children	59.8 %	69.6 %	74.3 %	70.1 %
Black or Latino	52.1 %	64.1 %	66.4 %	62.9 %
Not Black/Latino	66.5 %	74.6 %	80.4 %	76.1 %
Difference	-14.4 %	-10.5 %	-14.0 %	-13.2 %

We find overall a 15.8% drop in the proportion of parents who report their child's health as very good or excellent among those who have asthma, and almost twice that in younger children. Because 2 of our networks are in the greater NYC area, these data highlight children who live in more urban areas. Outside of urban areas both prevalence and gap between those with and without asthma are slightly higher (each ~17%). Effective delivery of guideline-based care can reduce the gap and decrease consequences of uncontrolled asthma, such as emergency room use and hospitalizations; better asthma care is beneficial and needed across the spectrum of children and primary care settings.[34-40] We find compelling evidence that the failure to effectively deliver guideline-based care contributes significantly to the lower health ratings for children with asthma, including for the 3.4 million urban Black and Hispanic children (age 2-17 years) with asthma. About 60% of these children are low income and have public insurance. We further are persuaded by evidence that quality of life and the quality of asthma management are associated specifically with such factors as family satisfaction with the nature of shared decision making.[41]

1c.4. Citations for data demonstrating high priority provided in 1a.3

33. ; Available from: www.childhealthdata.org.
34. Diette, G.B., et al., Consistency of care with national guidelines for children with asthma in managed care. The Journal of Pediatrics, 2001. 138(1): p. 59-64.
35. Adams, R.J., et al., Impact of Inhaled Antiinflammatory Therapy on Hospitalization and Emergency Department Visits for Children With Asthma. Pediatrics, 2001. 107(4): p. 706-711.
36. Finkelstein, J.A., et al., Underuse of controller medications among medicaid-insured children with asthma. Archives of Pediatrics & Adolescent Medicine, 2002. 156(6): p. 562-567.
37. Finkelstein, J.A., et al., Self-Reported Physician Practices for Children With Asthma: Are National Guidelines Followed? Pediatrics, 2000. 106(Supplement 3): p. 886-896.
38. Bell, L.M., et al., Electronic Health Record–Based Decision Support to Improve Asthma Care: A Cluster-Randomized Trial. Pediatrics, 2010. 125(4): p. e770-e777.
39. Lob, S.H., et al., Promoting Best-Care Practices in Childhood Asthma: Quality Improvement in Community Health Centers. Pediatrics, 2011. 128(1): p. 20-28.
40. Scott, L., et al., Achieving and maintaining asthma control in inner-city children. Journal of Allergy and Clinical Immunology, 2011. 128(1): p. 56-63.
41. Gandhi, P., et al., Exploring factors influencing asthma control and asthma-specific health-related quality of life among

children. Respiratory Research, 2013. 14(1): p. 1-10.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Prevention, Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Access, Care Coordination, Disparities, Overuse, Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
We currently do not have a web page. We will ensure that this measure will be publicly available.

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** FINAL_CAPQuaM_ASTHMA_ICD9_and_ICD10.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The numerator uses the number of undesirable utilization outcomes (i.e., claims for ED visits or hospitalizations for asthma) experienced by children who are managed for identifiable asthma to estimate the number of emergency room visits

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
This data requires 2 years of data, the reporting year and the 12 month period before the reporting year. (See Appendix 1, Figure 1)

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome

should be described in the calculation algorithm.

Numerator Elements:

Date and count of all emergency visits with a primary or secondary diagnosis of asthma.

ED visits should be identified as a visit that is associated with:

1) At least one of the following CPT codes: 99281, 99282, 99283, 99284, 99285 OR
2) At least one of the following revenue codes

0450 Emergency Room
0451 Emergency Room: EM/EMTALA
0452 Emergency Room: ER/ Beyond EMTALA
0456 Emergency Room: Urgent care
0459 Emergency Room: Other emergency room
450 Emergency Room
451 Emergency Room: EM/EMTALA
452 Emergency Room: ER/ Beyond EMTALA
456 Emergency Room: Urgent care
459 Emergency Room: Other emergency room
0981 Professional fees (096x) Emergency room
981 Professional fees emergency room

Inpatient Hospitalizations are identified as an encounter that is associated with:

At least one of the following CPT codes:

Hospitalization:

CPT 99238	CPT 99232
CPT 99239	CPT 99233
CPT 99221	CPT 99234
CPT 99222	CPT 99235
CPT 99223	CPT 99236
CPT 99356	CPT 99218
CPT 99357	CPT 99219
CPT 99231	CPT 99220

OR

At least one of the following revenue codes

0110	0133
0111	0134
0112	0137
0113	0139
0114	0150
0117	0151
0119	0152
0120	0153
0121	0154
0122	0157
0123	0159
0124	0200
0127	0201
0129	0202
0130	0203
0131	0204
0132	0206

IDENTIFY count of discrete numerator events:

For each individual in the denominator for the specified month, consider evidence of hospitalization that is on the same day or one day after an ED visit to represent one discrete event. Consecutive days of hospitalization are considered to represent one hospitalization.

Data Sources
Administrative Data (e.g., claims data)
Paper Medical Record – only if needed for race ethnicity or ZIP code
Race/ethnicity data and ZIP code data (If race/ethnicity data or ZIP code data are not present in administrative data set, they should be obtained from another source, such as the medical record). We performed a feasibility study alpha test by surveying more than a dozen hospitals that demonstrates that these data elements are generally available in the medical record.

General data elements:
- Age
- Race and ethnicity
- Insurance type (Medicaid, Private, Uninsured)
- Benefit type among insured (HMO, PPO, FFS, Medicaid Primary Care Case Management Plan [PCCM], Other)
- ZIP code or State and County of residence (and FIPS where available)

Administrative data with billing and diagnosis codes:
- Asthma-related visits to an emergency department, or hospitalization
- Asthma medication prescriptions
- Insurance benefit type
- ZIP code or State and County of residence (and FIPS where available)
- Race and ethnicity (from hospital administrative data or charts if not in administrative data from plan)

If pharmacy data are not available the measure should be reported with notation that pharmacy data were not used for the assessment of eligibility.

For eligibility purposes, asthma-related medicine refers to long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti-asthmatic combinations, methylxanthines (alone or in combination)

These details incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator represents the person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Populations at Risk, Populations at Risk : Individuals with multiple chronic conditions

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The denominator seeks to identify children who have been managed with identifiable asthma.

A descriptive definition for being managed for Identifiable asthma follows. Identifiable asthma needs to be identified in the assessment period for the specific reporting month being assessed.
Specifications follow the descriptive definitions:
a. Any prior hospitalization with asthma as primary or secondary diagnosis
b. Other qualifying events after the fifth birthday (age is age at occurrence):
i. One or more prior ambulatory visits with asthma as the primary diagnosis (this criterion implies an asthma ED visit in the reporting month), OR
ii. Two or more ambulatory visits with asthma as a diagnosis, OR
iii. One ambulatory visit with asthma as a diagnosis AND at least one asthma-related prescription, OR
iv. Two or more ambulatory visits with a diagnosis of bronchitis
c. Other qualifying events, any age:

v. Three or more ambulatory visits with diagnosis of asthma or bronchitis, OR
vi. Two or more ambulatory visits with a diagnosis of asthma and/or bronchitis AND one or more asthma- related prescriptions.
For eligibility purposes, asthma-related medicine means long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti-asthmatic combinations, methylxanthines (alone or in combination), and/or mast cell stabilizers.
If pharmacy data are not available, the measure should be reported with notation that pharmacy data were not used for the assessment of eligibility. This avoids eliminating from the measure those facilities with no link to pharmacies. Our testing reveals that only a very small proportion of patients are excluded by not including pharmacy data to establish eligibility.

For eligibility purposes, asthma-related medicine refers to long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti-asthmatic combinations, methylxanthines (alone or in combination), and or mast cell stabilizers. In order to promote better harmonization, we start with the current HEDIS asthma medication list. From that list, in accordance with our expert panel recommendations we eliminate medications in the following
2 categories: leukotriene modifiers, short-acting inhaled beta-agonists. We further exclude indacaterol, a recently approved long acting beta agonist that is indicated in the US only for teh treatmetn of COPD. As indicated elesewhere, COPD is an exclusion criterion for this measure. These specifications anticipate that NCQA will update the medication list from time to time and with the stated exclusions updated lists may be substituted for the list linked herein. The table used for testing is labeled Table AMR-A: Asthma Controller and Reliever Medications, and can be found at
http://www.ncqa.org/HEDISQualityMeasurement/HEDISMeasures/HEDIS2015/HEDIS2015NDCLicense/HEDIS2015FinalNDCLists.aspx (last accessed September 12, 2015).
Denominator Elements:

The presence of identifiable asthma (see Table 1) is established each month from administrative data using the specified algorithm. (Appendix Figure 1 and this section's narrative)

All events in the administrative data should be associated with a date of service.

Eligibility should be obtained using the month by month algorithm described herein and illustrated in Figure1, which is a fundamental component of this description. The analysis should be conducted on a month by month basis as described herein:
. Within the group of children who meet the criteria for identifiable asthma, identify and maintain a unique patient identifier, age, and all stratification variables.
. Determine eligibility for each patient, as of the last day of the month prior to the reporting month.

For example, if the goal is to report for January 2011, first identify children with identifiable asthma (above), and analyze all of calendar year 2010 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2010.

Next, for February analyze all of calendar year 2010 AND January 2011. Continuous enrollment criterion requires that the child was enrolled in December
2010 and January 2011.

Repeat this progression monthly so that for December, one would identify children with identifiable asthma and analyze all of calendar year 2010 AND January through November 2011 when doing so. Continuous enrollment criterion requires that for December the child was enrolled in October 2011 and November 2011.

See Figure 1 in Appendix, which is incorporated into these specifications by reference.

Codes used for definitions are specified in Appendix Table 1 and summarized herein:

Hospitalization:

CPT Codes: (Any)

CPT 99238 CPT 99232
CPT 99239 CPT 99233

CPT 99221 CPT 99234
CPT 99222 CPT 99235
CPT 99223 CPT 99236
CPT 99356 CPT 99218
CPT 99357 CPT 99219
CPT 99231 CPT 99220

Or Revenue Codes: (Any)

0110 0133
0111 0134
0112 0137
0113 0139
0114 0150
0117 0151
0119 0152
0120 0153
0121 0154
0122 0157
0123 0159
0124 0200
0127 0201
0129 0202
0130 0203
0131 0204
0132 0206

Emergency Department Visits

CPT Codes: (Any)

CPT 99281 CPT 99284
CPT 99282 CPT 99285
CPT 99283

Or Revenue Codes: (Any)

0450 Emergency Room
0451 Emergency Room: EM/EMTALA
0452 Emergency Room: ER/Beyond EMTALA
0456 Emergency Room: Urgent Care
0459 Emergency Room: Other Emergency Room
0981 Professional Fees (096x) Emergency Room
981 Professional Fees emergency room

Office Visits(Any)

CPT 99201 CPT 99211
CPT 99202 CPT 99212
CPT 99203 CPT 99213
CPT 99204 CPT 99214
CPT 99205 CPT 99215

Diagnosis of Asthma

ICD-9 Codes:
All codes beginning with 493
Alternately, or entities that prefer to use AHRQ's Clinical Classifications Software, the asthma definition before exclusions is CCS class 128. Those using CCS should then apply the exclusions.

Filled Prescriptions for Asthma-related Medications as specified in this section above.

Please note Figure 1 and Table 1 in the attached Appendix are considered INTEGRAL to these specifications and are not optional.

These details incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 Code: 496), Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03, 277.09), or Emphysema diagnosis (ICD-9 code 492xx).

These exclusion incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month and for the reporting month (a total of three consecutive months ending in the reporting month).

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
See S.10 above. Also, for entities that use AHRQ's Clinical Classifications Software, apply the exclusion after identifying visits that satisfy CCS class 128.

These details incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Specifications for this measure requires stratification by age group and race/ethnicity. Several additional stratifications are optional but may be required by the accountability entity or reported by the reporting entity. These variables include rurality/urbanicity and county level of poverty.

SPECIFICATIONS:

Identify County equivalent of child's residence. If County and State or FIPS code are not in the administrative data, the zip codes can be linked to County indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to County or County equivalents as used in various states.

i. Identify the Urban Influence Code (1) or UIC for the county of child's residence. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban-influence- codes.aspx#.UZUvG2cVoj8).

ii. Identify the Level of Poverty in the child's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using child's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:
a. Lowest Quartile of Poverty if percent in poverty is <=12.5%
b. Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%

c. Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
d. First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
e. Second Upper Quartile (>90th percentile)

iii. Categorize age by age at the last day of the month that ends the assessment period. Aggregate into age categories 2-4, ages 5 through 11, ages 12-18, ages 19-21.

iv. Categorize Race/Ethnicity as Hispanic, Non-Hispanic White, Non- Hispanic Black, Non-Hispanic Asian/Pacific Islander, and Non-Hispanic Other

v. Categorize Insurance Type as Private (Commercial), Public, None or Other

vi. Categorize benefit type as HMO, PPO, FFS, PCCM, or Other

NOTES on STRATIFICATION

Special Health Care Needs

The Maternal and Child Health Bureau has defined children with special health care needs (CSHN) as children "[w]ho have or are at increased risk for a chronic physical, developmental, behavioral, or emotional condition and who also require health and related services of a type or amount beyond that required by children generally". Considering this definition, children with Identifiable asthma are children with special health care needs.

Socioeconomic Status

The measure is specified to be stratified in 2 ways to assess aspects related to socioeconomic status: public versus commercial insurance and by 5 strata defined by the percent of the population in poverty in the county of residence. During our feasibility assessment phase, we asked more than a dozen respondent institutions whether the payment source was available in the medical record (EMR or paper) and the difficulty of abstracting this information from those records. We found that payment source is generally available in the medical chart and is overall not difficult to abstract. As we expect this measure primarily to be generated by insuring entities, these data are expected to be present and available in the administrative data. ZIP codes of residence are typically available in both medical records and administrative data sets and can be linked to county of residence as described in the specifications. We have identified five distinct strata based on the proportion of persons living beneath the poverty line. Such ecological data have been found to be independent predictors of health outcomes and are readily available using USDA data (1). The five strata represent the three quartiles of lowest poverty (each as one stratum) and the highest quartile divided into two strata, the 75th-90th percentiles and the highest 10%. In New York State only quartiles 1 through 3 are present, so we were not able to demonstrate the sensitivity of the measure specifically, but we were able to demonstrate the practicality of the method. Quartile one is the lowest proportion of households in poverty in the county.

Rurality/Urbanicity

These measure are specified to be reported by Urban Influence Codes (UIC), which have been developed by the USDA based on a number of criteria to describe the levels of urbanicity and rurality. This is intended not only to report within plan differences but to allow for aggregation as appropriate. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. We recommend consideration of the aggregation schema of Bennett and colleagues at the South Carolina Rural Research Center (2). Their aggregation scheme brings together Codes 1 & 2 as Urban; 3, 5, & 8 as micropolitan rural; 4, 6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We observe that UIC 5 might also be aggregated with 4, 6, & 7 as an adjacent rural area. While this approach to rurality does not map exactly to the population density based definition of frontier (< 6 persons per square mile) as articulated in the Affordable Care Act (ACA), use of such categories is consistent with the ACA's intent that the Secretary ask that data collected for racial and ethnic disparities also look at underserved frontier counties. Frontier health care may be approximated by analysis of the remote rural categories (3). Our judgment was confirmed after CAPQuaM consulted with Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Sciences, who is heading a HRSA-funded project to develop new methods to analyze frontier health. We clarified that his work suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in sensitivity.

Those interested in care specific to large cities may wish to aggregate rural areas and analyze UIC 1 and 2 separately. The New York

State Medicaid data was sensitive to urbanicity with higher rates of ED utilization in the most urban areas and lowest in the most rural areas and other areas intermediate between the two.

For aggregation and another approximation can also group as urban (1 and 2), suburban (3-6) and rural (7-9) [or 7-12 if more rural counties are present. NY State has not counties more rural than 9]. This is what we have used for our NY Medicaid analysis to demonstrate that variations are observed for this measure using UIC codes.

All of the above are specified and the choice of schema (if any) belongs to the accountability entity.

1.Kawachi I, B.L., Neighborhoods and Health. 2003, New York, NY: Oxford University Press.
2.Bennett, K.J., Olatosi B. & Probst, J.C., Health Disparities: A rural-urban chartbook. 2008, Columbia, South Carolina: South Carolina Rural Health Research Center.
3.Hart, G., Frontier/Remote, Island, and Rural Literature Review. 2012.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Other
If other: In order to allow for more granular comparisons this measure is specified to be stratified. Stratification for risk adjustment of this measure would not be justified by the literature. Although epidemiological findings support our stratification schema, no biological evidence exists to support intrinsic correlation of ED rates with stratification variables.

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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.)*
Step 1: Measure person-time eligible for each patient and record by month.
a. For each month in the reporting year, identify all children ages 2 – 21 years who meet the criteria for Identifiable asthma during the assessment period. The assessment period is defined as the year prior to the reporting year plus all months in the reporting year prior to the reporting month.

Identify and maintain a unique patient identifier and all stratification variables.

To illustrate: if the goal is to report for January 2011, first one would identify children with Identifiable asthma using the criteria, and analyze all of calendar year 2010 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2010, as well as January 2011. This total represents the number of person-months (child-months) for January.

Next, for February: one would identify children with Identifiable asthma using the criteria, and analyze all of calendar year 2010 AND January 2011 when doing so. Continuous enrollment criterion requires that the child was enrolled in December 2010 and January 2011, as well as February 2011. This is the number of person-months (child-months) for February. Repeat this progression monthly so that for December, one would identify children with Identifiable asthma and analyze all of calendar year 2010 AND January through November 2011 when doing so. Continuous enrollment criterion requires that the child was enrolled in October 2011 and November 2011, as well as December 2011. This is the number of person-months (child-months) for December.

b. Sum all months that are eligible from the reporting year. This sum is the denominator in people-months. Divide by 1200. This is denominator in 100 people-years. This is the denominator for the year.

Step 2: Month by month, considering the definitions above, identify the number of discrete numerator events:
a. Identify the number and date of ED visits with asthma as a primary or secondary diagnosis among those children who are eligible for that reporting month.
b. Identify the number and date of inpatient hospitalizations with asthma as a primary or secondary diagnosis among those children who are eligible for that reporting month.
c. Identify the number of discrete numerator events. Consecutive days with inpatient hospital codes are considered one hospitalization. Hospitalizations on day of or day after ED visit are NOT considered discrete from the ED visit.
d. Sum the number of numerator events across the year.
e. Maintain stratification variables and unique identifiers.

Step 3. Calculate rate as Numerator / Denominator. While this measure is specified for the year, it has also been validated to demonstrate seasonality using monthly rates.

Step 4. Calculate stratification variables as specified in S.12.

Step 5. Repeat by strata. Within age strata repeat by other specified strata. Perform other cross tabulations as requested by the accountability entity. Eliminate any strata with less than 40 person-months in any month's denominator OR less than 1000 person-months for the year.

Appendix 1, Figure A.1 illustrates the calculation of person-time and is considered fundamental to this calculation algorithm.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
When data cannot be obtained from any source:
If critical for calculation – delete patient from consideration for that reporting month
If non-critical for calculation – include patient
Critical data include encounter data for the reporting month and some period of time in the assessment period. In order to report stratifications age and race/ethnicity are considered critical.

Pharmacy data are not considered critical

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
N/A

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System, Population : Community, Population : County or City, Population : National, Population : Regional, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Ambulatory Care : Urgent Care, Emergency Medical Services/Ambulance, Hospital/Acute Care Facility, Other, Pharmacy
If other: Claims data from all settings in New York State Medicaid data were tested.

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Testing_Attachment_Asthma_1_Round_2_Final.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: A PQMP Measure
Date of Submission: 1/6/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be

completed.

- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record (limited data elements)	☒ abstracted from paper record (limited data elements)
☒ administrative claims	☒ administrative claims

☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record (limited data elements)	☒ abstracted from electronic health record (limited data elements)
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

New York State Medicaid claims data 2010 – 2012.

Also, our work builds off of work performed by our CAPQuaM partner and steering committee member, NCQA. For specific data reliability and signal to noise analyses, we incorporate by reference (and will present more selectively) NCQA data relevant to their submission for NQF –endorsed asthma related measures:

- Use of Appropriate Medications for People with Asthma (ASM) – 0036 (we understand this is no longer being maintained as of 2015, but it was endorsed and the data were accepted.)
- Medication Management for People With Asthma (MMA) – 1799
- Asthma Medication Ratio (AMR) – 1800

We note that 1799 and 1800 are not directly applicable because they were tested at the score level. However, the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to produce valid scores.

The analyses above provide information regarding the capacity to use administrative data to identify the applicable denominator population. There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures.

1.3. What are the dates of the data used in testing? 2010-2012

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☒ other: Population, State, Region, County, Integrated delivery system	☒ other: Population, State, Region, County, Integrated delivery system

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

NY State Medicaid Managed Care claims data, including claims from all MCO's that are contracted for Medicaid care by our partner, the NY State Department of Health. The numbers we present are from reporting year 2012, include children from counties in nine urban influence codes and in counties poverty level 1-3. NY State does not have any counties in the lowest 25% of poverty or with UIC of 10-12. New York has more than 60 counties and numerous health plan vendors. Analysis in Year 2011 provided very similar data.

Analysis of NY State Medicaid Managed Care claims data, including claims from all MCO's that are contracted for Medicaid care by our partner, the NY State Department of Health. We identified eligible populations and events from both Reporting Year 2011 and 2012 and include children from counties in nine urban influence codes and in counties poverty level 1-3. NY State does not have any counties in the lowest 25% of poverty or with UIC of 10-12. New York has more than 60 counties and numerous health plan vendors. Analysis in year 2011 provided very similar data to 2012.

Foundational analyses for this measure were performed and previously reported by NCQA considering *nine health plans covering a variety of geographic areas within the United States that were asked to provide a complete administrative data file consisting of any member in their commercial and Medicaid product lines for anyone that had a diagnosis code for asthma during the calendar years of 2009-2010. The complete member-level administrative file used for analysis included a total of more than 82,000 health plan members with asthma*.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

All children 0-21 with records in the 2011 in the 2011 NYS Medicaid Managed Care administrative database and all pediatric patients meeting the criteria for identifiable asthma in the 2012 NYS Medicaid Managed Care administrative database. These children experience 41,339 qualifying emergency department visits in that time frame of which 38,566 were in children 2-18. In 2011, the median number of visits per child with an ED visit was 1, the 75th percentile was 2, and the 90th percentile was 3.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data source 1 (Chart):
Prior to initial specification of the measure we contracted for a survey of quality managers representing more than a dozen hospitals to assess data availability and the ease and feasibility of abstraction of data relevant for asthma and other CAPQuaM measure, including 10 hospitals that responded regarding asthma-specific data elements. Our survey found that availability in the medical record of age, (date of birth), race, ethnicity, date and site of visit, documentation of primary or secondary diagnosis of asthma, hospitalization, and payment source were routinely available in the chart and "Not Difficult to Collect." Our chart review of 1200 medical records for ED visits in a single institution performed for validation of a sister measure on appropriateness confirmed the availability of these data. This validates the capacity to obtain such data form the medical record and the primary occurrence of the data in the chart so that coders have the clinical information required to population ICD-9/10, CPT, and Revenue codes that comprise the administrative data that are the preferred data source for these measures.

Data Source 2 (Administrative):
Assessment of the capacity to identify the eligible population and qualifying events was performed in NY State Medicaid data in both 2011 and 2012 reporting years.

Data source 3 (explicit criteria):
Our construct for the CAPQuaM measure was defined by the multidisciplinary national expert panel using a RAND type modified Delphi process, which produced a set of explicit criteria that were both substantive and addressed specification details, such as what combination of administrative codes could be used to identify a child with asthma, other inclusion and exclusion criteria for the measure, and preferences regarding how to report and stratify the measure.

The panel initially used the term persistent asthma to describe asthma that was pre-existing and should have been recognized as asthma by the health care system prior to the timing of the ED visit. This construct was renamed by our stakeholder group to be identifiable asthma to avoid confusion with other uses of the term persistent asthma. The construct was intended to be more inclusive than HEDIS' persistent asthma diagnosis, while still removing from consideration those whose asthma was unlikely to have been actively managed at the time.

Data Source 4 (National Survey of Children's Health)
We validate the construct of identifiable asthma comparing it to two other constructs:
HEDIS' definition of persistent asthma, which should have been more restrictive than 'identifiable asthma'; and the National Survey of Children's Health's question regarding if the caregiver had ever been told by a doctor or nurse that the child had asthma, which should have been less restrictive than 'identifiable asthma.' The former analysis was conducted in Medicaid 2011 and the latter in the most recent NSCH data.

Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1% with the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. As expected, identifiable asthma was between 2 and 3 times more permissive than the intentionally restrictive persistent asthma. We analyzed NSCH data to estimate a population rate of asthma in NY State Medicaid child population to be between 15 - 16%, indicating that our criteria did provide a meaningful filter as we had intended.

Reducing the continuous enrollment period down to three months as was suggested by members of our steering committee increases the number of children eligible for the measure by several tens of thousands while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group. This inclusiveness help to counter risks of churning that are particularly prominent in the Medicaid population. This analysis was conducted in the NYS Medicaid data.

Data Source 5: HEDIS
Assessment of data elements for identifying a population with asthma and asthma scores was performed by NCQA in nine geographically diverse managed care plans. We considered the HEDIS data for measures 1799, 1800 and 0036. We cite 1799 and 1800 not as specific evidence of score level performance of our measure, but as evidence that measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal from noise at a very high level.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Race, ethnicity, zip code, level of poverty in the zip code of caregiver residence, and urban influence in the county of caregiver residence for the NY State analysis. Within the Medicaid data, we looked at eligibility category.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Validity testing was performed at the data element level for both the numerator and the denominator. See section 2b2 for validity testing of data elements.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**
☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
Variation associated with stratification variables was similar whether using the 2011 or the 2012 data from New York State, suggesting test, retest reliability. Similarly, month-to-month variations in the two reporting years assessed followed the same patterns. That these patterns were identified and similar in two consecutive years of analysis, indicates reliability. Further the identification of seasonal changes within a defined population is a more difficult challenge for distinction (or signal to noise) than comparison in distinct populations.

Please see descriptions of both NCQA and CAPQuaM testing above in 1.2-1.7.

As described in the next section the literature also supports the use of claims data to identify the presence of asthma.

We develop our measure using scientifically sound principles. We first discuss research involving the soundness of our data sources, which include both administrative data to identify cases (and a fraction of numerator qualifications) and chart review (medical record audit) to confirm some denominator inclusions and to identify most numerator inclusion. This is a generally accepted and standard approach with acceptable reliability.

We use administrative data to identify the age of the child, various stratification variables and the presence of asthma, as well as the presence of an asthma ED visit or hospitalization. These are routinely used to support billing by CMS, Medicaid, and private insurers and are routinely used in quality measurement. Administrative data are not typically sufficient for detailed clinical assessment.[1-5] HEDIS developed a hybrid approach, using administrative data and chart review that this measure borrows heavily from. [6, 7]

There is moderate agreement (kappa = 0.45 – 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data. The agreement improves from 0.55 to 0.60 when using two years of data. (8). We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

1. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.
2. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p.

   199-207.
3. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p.

   813-25.
4. Angier, H., et al., *Variation in outcomes of quality measurement by data source.*

   Pediatrics, 2014. **133**(6): p. e1676-82.
5. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.
6. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.
7. NCQA. *National Committee for Quality Assurance*. [cited 2014 7/30/14]; Available from: http://www.ncqa.org/
8. Huzel, L, et al. Diagnosing Asthma: The fit between survey and administrative database. Can Respir J. 2002 Nov-Dec;9(6):407-12.

Table 1. 360 Degree Pediatric Quality Measure Development: Overview

Stage	Phase	Innovation	Product(s)
1. Clinical Criteria Development	a. Input Development	1. Focus groups of caregivers of children with asthma who have used the ED 2. Interviews with front line clinicians: primary care, asthma docs, and ED docs	1. Literature review 2. Summary of consumer perspectives, values and understanding relevant to clinical issue of interest 3. Summary of findings form clinician interviews
	b. RAND/UCLA 2 Round Modified Delphi Process	1. Inclusion of consumer perspectives as a key input; 2. Use of this method to identify appropriateness criteria in national performance measure development;	1. Explicit criteria that rank a comprehensive and mutually exclusive set of clinically detailed scenarios;

2. Boundary Guideline Development	Criteria Enhancement	1. Iterative process to enhance reliability and internal consistency of the explicit criteria set with a goal of outlining three boundary spaces	1. Internally consistent set of explicit criteria that are stable in their representation of the expert panel perspective. "Enhanced criteria"
	Guideline Articulation	1. Stakeholder (including experts, users, clinicians, consumers and others) informed review of the enhanced criteria. 2. Definition of zones of potential overuse, potential underuse, and professional interaction and decision-making based upon the explicit criteria 3. Stakeholder valuations of potential deviations from guideline 4. Boundary Guideline	1. Boundary Guideline 2. Prioritization list
3. Creation of Measure	Specification	1. Translation of guideline into specification of necessary data 2. Iterative process to define optimally efficient sources of data to allow for measurement and stratification	1. Initial specification of measure
	Review	1. Constructive peer review of specifications by stakeholders in Steering Committee and SAB	1. Final specifications of measure including variables for stratification as needed
	Fielding and testing of measure	1. Measure testing	1. Functional experience and practical understanding of measure, its scoring, variability, and interpretation

This measure was developed and assessed using a pre-specified process and consistent with CAPQuaM's peer reviewed 360 degree method outlined in the table above.

Explicit criteria were developed using a variation of the two-round modified Delphi process RAND/UCLA Appropriateness Method with a multidisciplinary and geographically diverse expert panel comprised of both clinicians and researchers. Identifiable asthma was based on panel findings and appropriateness criteria included for this measure were those that were both available in the chart and highly rated. The general reliability of this approach is well established. [9, 10] It has been applied successfully to pediatric services previously. [11-13]

9. Fitch, K., et al., *The RAND/UCLA Appropriateness Method User's Manual*. 2001 RAND.

10. Kosecoff, J., et al., *The appropriateness of using a medical procedure. Is information in the medical record valid?* Med Care, 1987. **25**(3): p. 196-201.

11. Kleinman, L.C., et al., *The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States.* Jama, 1994. **271**(16): p. 1250-5.

12. Kleinman, L.C., E.A. Boyd, and J.C. Heritage, *Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians.* Jama, 1997. **278**(6): p. 497-501.

13. Keyhani, S., et al., *Overuse of tympanostomy tubes in New York metropolitan area: evidence from five hospital cohort.* Bmj, 2008. **337**: p. a1607.

Development included a series of alpha tests to refine specifications by conducting iterative analyses in New York State Medicaid data. Conclusions from alpha tests include:

1) The reporting period and the assessment period could not overlap completely, leading to use of 2 years of data as shown in the specifications' diagram. The optimal approach was to divide the reporting year into 12 reporting months. ED events in that month are eligible for the numerator if persistent asthma criteria have been satisfied (combining the look-back year and all prior months in the reporting year) and the child has been continuously enrolled for the two months immediately prior to the reporting month. The optimal building block unit for the denominator is in child-months, which is rolled up to child-years;

2) Using both revenue codes and CPT codes increased our sensitivity meaningfully, a choice validated by consultation with coding and billing experts and confirmed by analyzing the NY State data;
3) NY State Medicaid data and national survey data (HCUP) converged to demonstrate the importance of including hospitalizations as numerator events even when the underlying construct is ED visits. This is consistent with policies of many payers to request providers not to submit both ED and hospital claims for the same day. Error is far less by considering both ED visits and hospitalizations as numerator events, than by not including hospitalizations.
4) The expert panel only wanted numerator events for which the children were already known to the accountable entity as having asthma and established definitions for such "identifiable asthma". Alpha testing in NY State Medicaid demonstrated the expected results:
    a. Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1%, the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. This more inclusive approach was our goal.
    b. More than 25% of children with any asthma claim are not included in the denominator, indicating that this is a meaningful filter. Confirming this, the observed rate of 8.3% in the denominator. 8.3 is just over half of what we found when analyzing NSCH data to identify an expected rate of NY State Medicaid children whose caregivers would report that they every been told the child had asthma.
    c. Relaxing the continuous enrollment period to 3 months was suggested by members of our stakeholder steering committee. Doing so increased the eligible number by more than 20,000 while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children. [14]

14. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.

Key panel ratings are shown. Constructs rated 7 or higher are endorsed, 8 or higher strongly endorsed, and 2 or lower strongly rejected.

The definitions were specified to allow their use with data elements that ought to be available in electronic form to a responsible entity, such as a health plan or state Medicaid program. Potential exceptions to this are elements such as ZIP code of residence and race and ethnicity of the child. We have data from a feasibility study we conducted with a contractor that surveyed quality departments at more than a dozen hospitals across three measure sets. 10 hospitals responded to the asthma-specific questionnaire. We found that these data elements are generally available in the chart, although the definition of race and ethnicity, as well as how it is determined, may vary by institution. Nonetheless, the CHIPRA legislation (2009), which has funded the development of this measure, directs for measures to be capable of identifying disparities and we have specified it to be so, despite concerns about reliability in the collection and assessment of race and ethnicity by health-care-providing institutions and practices. In this case, we need to drive performance through measurement, as it is foundational to the legislative and executive branch sources of our funding.

This process has led us to enhance the validity of this measure by deflating competing concepts and clearly specifying it as an interpretable epidemiological rate (incidence density). The former Medicaid core measure that we were tasked with enhancing was a simple risk, with asthma patients defined in the measurement year as having primary or secondary diagnosis for any service, and ED visits defined as CPT-code-identified ED visits with asthma as the primary diagnosis. The numerator for the Core Measure includes all patients with at least one ED visit for asthma as asthmatic events, whether or not the patient was known to be an asthmatic before the event. Further, numerator events alone could qualify children for inclusion in the denominator. Our partners

in the New York State Medicaid program have described this characteristic as highly undesirable and the CAPQuaM team agreed, prompting our month by month approach to analysis.

The name for this measure is "Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma" is telling and embodies a number of ways we attempted to enhance the validity of the measure: one episode of asthma or asthma-like systems will not necessarily qualify a child as having identifiable asthma; identifiable asthma must precede the asthma visit; the child must have received some treatment for services that suggest asthma, thus making the fact that the child has asthma available to the health care system; and we are considering a rate and not a risk: more than one visit to the ED from each child results in counting each distinct visit. While the median number of visits among those with visits is 1, more than one-quarter of children in New York State Medicaid Managed Care with an ED visit have a second visit. A few outliers contribute more than 10 Ed visits per child. The rate measure allows us to provide a better estimate of the number of undesirable outcomes, rather than the number of children with undesirable outcomes.

As a rate, one child can contribute to the numerator many times. It also is self-adjusting for children who enter or leave the eligible population since children contribute to the denominator independently for each month that they are eligible. It also assures that ages can be calculated to the month rather than to the year. Further, in an attempt to enhance the meaningfulness of the measure, we have included a two-month continuous enrollment requirement prior to the reporting month. Since the child must also be eligible for the reporting month, this becomes a three-month continuous enrollment requirement. In doing this, we sought to strike a delicate balance between developing a meaningful accountability measure and eliminating children because of problems of churning, which have been well documented by researchers (15). This balance was achieved in close collaboration with our colleagues at NY State Medicaid.

We pre-tested our specifications with a series of iterative analyses in New York State Medicaid data. Early on, we found that the combined definitions of persistent asthma and the need for the diagnosis to precede the ED visit meant that the reporting period and the assessment period could not overlap completely. These tests led us to analyze two years of data, as shown in the diagram included with our specifications; one year is the reporting year and one the look-back year. We further divide the reporting year into 12 reporting months. ED events in that month are eligible for the numerator if persistent asthma criteria have been satisfied (combining the look-back year and all prior months in the reporting year) and the child has been continuously enrolled for the two months immediately prior to the reporting month.

We also found many visits in 2011 that were identified by revenue codes and not by CPT codes; using both increased our yield substantially. After consultation with several coding and billing experts and reviewing existing approaches to measurement, we incorporated revenue codes into our specifications.

Identifiable asthma was defined according to the results of an Expert Panel that was intending to specify a subset of children that would be more inclusive than currently existing approaches, such as the HEDIS Hospitalizations for Children with Persistent Asthma measure. Testing demonstrated the expected results.

Because many health plans including New York State Medicaid encourage providers to NOT submit claims for ED visits that are associated with hospital admissions, there is the potential for systematic under identification of ED visits. Analysis of both NYS Medicaid data and national data (see poster that follows in the next section) confirms that the introduction of error would be far less by considering both ED visits and hospitalizations as numerator events, than by not including hospitalizations. To reduce the possibility of double counting, we de-duplicate admissions that are the day of or day following an ED visit. While this may introduce undercounting if some back to back ED visits result in admission from the second visit, we expect the error to be small. Since we view this as a clinical outcome and since the episode of care would be associated with a numerator event, we do not feel this threatens validity in a meaningful way. The exact number is not readily assessed (if the data were perfect, there would be no need for this aspect of our protocol).

The development team's goal was to develop an ICD10 code set that was fully consistent with the intent of the original measure. Our process began by performing general equivalency mapping using the forward mapping from www.icd9data.com. We then did a de novo review of the CMS ICD 10 CM set to seek to identify codes that might be appropriate for asthma. We reviewed potential codes identified by both sources and developed a new list of codes appropriate for inclusion criteria and a new list of codes appropriate for exclusion criteria. Drs. Kleinman and Sharma reviewed the lists independently and then achieved consensus in a conference call review and discussion. Key team members for this work were Suzanne Lo, MPH who staffed and coordinated this work, Sandeep Sharma, MD, Dr.PH and Lawrence Kleinman, MD, MPH. Dr. Sharma was a lead developer for one of CAPQuaM's 2 asthma measures and Dr. Kleinman is both CAPQuaM PI and was a lead developer for both measures. The guidance for the intended constructs for both ICD9 and ICD10 coding were the findings from a RAND style modified Delphi panel that incorporated 9 national experts over the course of the measure development process.

15. Fairbrother, G., et al., Churning in Medicaid managed care and its effect on accountability. J Health Care Poor Underserved, 2004. 15(1): p. 30-41.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
IDENTIFYING A POPULATION WITH ASTHMA:

For the foundational NCQA work (Measures 1799, 1800, 0036), NCQA's field test retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. Using the dataset provided, NCQA examined several different scenarios to determine the effects of different specification criteria on this particular population. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims.

From NCQA's submissions: *Reliability was estimated on the HEDIS 2011 submissions (2010 data) using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.*

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

We note that 1799 and 1800 are not directly applicable because they were tested at the score level. However, the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to produce valid scores.

Thus cite them not as specific evidence of our score level performance of the submitted measure, but as evidence that the HEDIS measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal to noise at a very high level. If the population assessment were inadequate, then these other measures which use the same data elements to establish their denominators could not achieve such high reliability scores. This is because failure to distinguish signal from noise at the level of the HEDIS denominators would lead to non-differential misclassification error which is a major bias towards the null, in other towards noise and away from signal. Hence these provide strong indirect evidence of the validity of our approach to capturing the measure's denominator.

The analyses above provide information regarding the capacity to use administrative data to identify the applicable denominator population. There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures.

Further, we identify asthma visits and medications using the same data that an insurance company or Medicaid would use for payment, including ICD9 codes, CPT codes, and revenue codes. We have had conversations with

expert coders and New York State Department of Health Office of Health Insurance Programs to confirm our choices.

Review of the medication lists for 0036 reveal that all medication used by the submitted CAPQuaM measure are also in the HEDIS measure. The CAPQuaM measure excludes specifically short acting beta agonists and leukotriene inhibitors at eh specific direction of the CAPQuaM expert panel. We also specify to exclude indacaterol, a long acting beta agonist which is only indicated in the USA for treatment of COPD, which is a specific exclusion criterion for this measure.

Our literature review found that while there is moderate agreement (kappa = 0.45 - 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data to parent report, the agreement improves from 0.55 to 0.60 when using two years of data.(1) We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.
The literature further supports our work. ICD-9 and ICD-10 codes for asthma on patients' medical charts typically match claims data. ICD-9-CM administrative data have been validated using various methodologies for various purposes (2-10). As examples: Jollis et. al. compared insurance claims data to the clinical database data to identify patients using ICD-9-CM codes for selected diagnoses and found that when all diagnoses were included, overall kappa agreement was .75 (2). Lee et. al. compared heart failure diagnoses identified in ICD-9 to the Framingham clinical criteria as the gold standard and found a positive predictive value of 94.3% (3). Muhajarine et. al. compared self-reported heart health survey data to physician claims from a database registry and found an overall agreement for hypertension of 81.7% indicating moderate to high agreement(4).Quan et. al. tested administrative discharge data to chart data for recording of comorbidity information using a Charlson index for measurement. Overall agreement of the Charlson index was good between databases but decreased as burden of comorbidity increased. Despite the differences, the Charlson index score derived from the administrative data had an identical ability of predicting in-hospital mortality to the score derived from chart data (5). Weiner and colleagues advocate a broad use of admisntrative data for monitoring quality and our uses fall within their recommendations (6). Romano and Mark assessed the sensitivity and reliability of coding for common diagnoses and procedures using California discharge abstracts and found in 7 of 8 comorbidity categories, sensitivity exceeded 85% (7). Weingart et. al. used administrative data, specifically a complications screening algorithm to identify inpatient complications using physician judgment as the gold standard and found flagged complications in 68.4% of surgical cases and 27.2% of medical cases (8). Yasmeen et. al. examined the sensitivity and positive predictive value to validate the coding of obstetric diagnoses and procedures in hospital-reported data using the medical record as the gold standard and found that surgical procedures and birth deliveries were accurately reported with sensitivities and PPVs exceeding 90% (9). Quam et.al. found that claims data that includes diagnostic and pharmacy data yields a high level of concordance with the medical record and survey data in the identification of a specific medical condition (10). Studies have shown high sensitivity of 72% and specificity of 95% for high risk conditions with overall accuracy of 90% obtained from administrative billing data among children with high-risk conditions including asthma which made up 87% of the high risk conditions (11), and high predictive value among adolescents and adults with asthma (12). Twiggs et. al. found that the combined use of both medical and pharmaceutical claims was more effective in identifying asthmatics than either one by itself (13). HEDIS criteria using administrative data support peer reviewed research, for example in patients with persistent asthma based on HEDIS criteria in five Medicaid programs (Colorado, Georgia, Indiana, New Jersey, Washington) using ICD-9-CM code 493.x to measure filling prescriptions of asthma control medication and the ratio of controller medication to the total number of medication prescriptions filled within one year (14). Fowles and colleagues report sensitivity and specificity of claims compared with ambulatory medical records to identify asthma was 0.82 and 0.99, respectively. Sensitivity of .82 using claims was higher than sensitivity using self-report at .64 (15). Wilchesky compared chart abstraction to diagnoses obtained from administrative database: asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0). Although sensitivity for most conditions was below 60%, sensitivity was enhanced when all claims for services were assessed, as we propose to do (16). Bronstein et al found that 88.3% of

diagnoses asthma on claims agreed with medical record, with a negative predictive value of 0.85 and a positive predictive value of 0.88.They conclude that claims are generally an accurate indicator of the content of a patient encounter. (17) Steinwachs et al. compared billed claims to medical records based on date of visit and diagnosis, on average, 90% of billed visits were documented in the medical record, for asthma there was 90.9 percent of billed visits in record on same date and 82.8 percent of billed visits with same diagnosis in record on same date. (18) Quan et al documented the validity of ICD-9-CM and ICD-10 coding systems in coding clinical information and found that ICD-10 data was generally comparable with that of ICD-9-CM data in recording clinical information (19). Regarding our capacity to identify exclusions, Quan et al found that claims had a PPV of 91.9, and a negative predictive value of 92.6, with *k* of 0.65 (substantial agreement[1]) compared to chart review for chronic pulmonary disease . ICD 10 performed similarly in this study (19).

From a public health perspective, asthma surveillance systems in several states, including Maine, North Carolina, Connecticut and Michigan, have shown the feasibility of using administrative data to identify children having asthma, based on primary and secondary diagnosis codes reported on inpatient and outpatient claims. In addition to identifying asthma, important demographic data such as gender, race/ethnicity, program of enrollment and county of residence (urbanicity) can be used to assess associations between utilization services for asthma, including ED visits or hospitalizations, and demographic characteristics. Risk factor information from administrative data can be used to target educational programs, clinical assessments, and treatment programs (20-23).

Researchers also classified children with evidence of persistent asthma using HEDIS criteria, (24). Another study showed the usefulness of ICD9 493.x to identify asthma for a quality measure using Maryland Medicaid Claims data (25). Like our measure, those researchers excluded children with a diagnosis of cystic fibrosis (ICD9 277) (25). Schneeweiss commented that misclassification errors from claims data are asymmetric, with specificity typically exceeding 95% and sensitivity often less (26). Such a pattern makes it unlikely that an accountable entity would be held accountable for patients that do not actually have asthma.

As noted in 1.67 above, as part of an alpha test for our measure we used a contractor to survey more than a dozen hospitals across three CAPQuaM measure sets. Responses from 10 hospitals were specific to asthma. We found that variables including date of birth, race, ethnicity, county of residence, primary and secondary diagnosis of asthma in the ED, hospitalizations, payment source, and others were reported to be readily available and easy to within the medical record.

In light of the literature review and our alpha test, we attest that the data elements for the measure match those assessed in the literature and our alpha test, with most being supported by both the literature review and the alpha test.

REFERENCES

(1) Huzel, L. et al. Diagnosing Asthma: The fit between survey and administrative database. Canada Resp. Journal 2002.
(2) Jollis, J.G., M. Ancukiewicz, E.R. DeLong, D.B. Pryor, L.H. Muhlbaier, and D.B. Mark. 1993. ‘‘Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research.’’ Annals of Internal Medicine 119: 844–50.
(3) Lee, D.S., L. Donovan, P.C. Austin, Y. Gong, P.P. Liu, J.L. Rouleau, and J.V. Tu. 2005. ‘‘Comparison of Coding of Heart Failure and Comorbidities in Administrative and Clinical Data for Use in Outcomes Research.’’ Medical Care 43: 182–8.
(4) Muhajarine, N., C. Mustard, L.L. Roos, T.K. Young, and D.E. Gelskey. 1997. ‘‘Comparison of Survey and Physician Claims Data for Detecting Hypertension.’’ Journal of Clinical Epidemiology 50: 711–8.
(5) Quan, H., G. Parson, and W. Ghali. 2002. ‘‘Validity of Information on Comorbidity Derived from ICD-9-CM Administrative Data.’’ Medical Care 40: 675–85.
(6) Weiner, J.P., Powe,. R., Steinwachs, D.M., & Dent, G. (1990). Applying insurance claims data to assess quality of care: A compilation of potential indicators. Quality Review Bulletin, 16, 424-438.

[1] The *k* value indicates a near perfect agreement (*k*: 0.81-1.0 between coded data and chart review data), substantial agreement (*k*: 0.61-0.80), moderate agreement (*k*: 0.41-0.60), and fair agreement (*k:* 0.21-0.40).

(7) Romano, P.S., and D.H. Mark. 1994. ‘‘Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment.’’ Medical Care 32: 81–90.
(8) Weingart, S.N., L.I. Iezzoni, R.B. Davis, R.H. Palmer, M. Cahalane, M.B. Hamel, K. Mukamal, R.S. Phillips, D.T. Davies Jr., and N.J. Banks. 2000. ‘‘Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program.’’ Medical Care 38: 796–806.
(9) Yasmeen, S., P.S. Romano, M.E. Schembri, J.M. Keyzer, and W.M. Gilbert. 2006. ‘‘Accuracy of Obstetric Diagnoses and Procedures in Hospital Discharge Data.’’ American Journal of Obstetrics and Gynecology 194: 992–1001.
(10) Quam, L., Ellis, L. B. M., Venus, P., Clouse, J., Taylor, C. G., & Leatherman, S. (1993). Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. Medical Care, 31(6), 498-507.
(11) Daley MF, Barrow J, Pearson K, Crane LA, Gao D, Stevenson JM, et al. Identification and recall of children with chronic medical conditions for influenza vaccination. Pediatrics 2004;113(1 Pt 1):e26-33.
(12) Vollmer WM, O’Connor EA, Heumann M, Frazier EA, Breen V, Villnave J, Buist AS. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol 2004;57:392-7.
(13) Twiggs, J. E., Fifield, J., Apter, A. J., Jackson, E. A., & Cushman, R. A. (2002). Stratifying medical and pharmaceutical administrative claims as a method to identify pediatric asthma patients in a Medicaid managed care organization. Journal of Clinical Epidemiology, 55(9), 938-944.
(14) Samnaliev, M., Baxter, J. D., & Clark, R. E. (2009). Comparative evaluation of two asthma care quality measures among Medicaid beneficiaries. Chest, 135(5), 1193-1196.
(15) Fowles, J. B., Fowler, E. J., & Craft, C. (1998). Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. Journal of Ambulatory Care Management, 21(1), 24-34.
(16) Wilchesky, M., Tamblyna, R. M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. Journal of Clinical Epidemiology, 57, 131-141.
(17) Bronstein, J. M., Santer, L., & Johnson, V. (2000). The use of Medicaid claims as a supplementary source of information on quality of asthma care. Journal for Healthcare Quality, 22(6), 13-18.
(18) Steinwachs, D. M., Stuart, M. E., Scholle, S., Starfield, B., Fox, M. H., & Weiner, J. P. (1998). A comparison of ambulatory Medicaid claims to medical records: A reliability assessment. American Journal of Medical Quality, 13(2), 63-69.
(19) Quan, H., Li, B., Saunders, L. D., Parsons, G. A., Nilsson, C. I., Alibhai, A., et al. (2008). Assessing validity of icd-9-cm and icd-10 administrative data in recording clinical conditions in a unique duallycoded database. HSR: Health Services Research, 43(4), 1424.
(20) Deprez RD, Asdigian NL, Oliver LC, Anderson N, Caldwell E, Baggott LA. Development of a prototype system for statewide asthma surveillance. Am J Public Health 2002;92:1946-51.
(21) Buescher PA, Jones-Vessey K. Using Medicaid data to estimate state- and county-level prevalence of asthma among low-income children. Matern Child Health J 1999;3:211-6.
(22) Children’s Health Council. Asthma affects nearly one in ten children enrolled in HUSKY Part A in 1999–2000. Hartford (CT): Children’s Health Council; 2001.
(23) Dombkowski, K. J., Wasilevich, E. A., & Lyon-Callo, S. K. (2005). Pediatric asthma surveillance: Using Medicaid claims. Public Health Reports, 120, 515.
(24) National Committee for Quality Assurance. Use of appropriate medications for people with asthma. HEDIS 2003, volume 2: technical specifications. Washington: NCQA; 2003.
(25) Stuart, B., Singhal, P. K., Magder, L. S., & Zuckerman, I. H. (2003). How robust are health plan quality indicators to data loss? A monte carlo simulation study of pediatric asthma treatment. HSR: Health Services Research, 38(6), 1547.
(26) Schneeweiss, S., & Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. Journal of Clinical Epidemiology, 58(4), 323-337.

For the foundational NCQA work, NCQA’s field test retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. Using the dataset provided, NCQA examined several different scenarios to determine the effects of different specification criteria on this particular population. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims.

They report that score level reliability of the HEDIS 2011 submissions (2010 data) was assessed using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

Our own research looking at NY State Medicaid and national all payer data (see poster below, which was presented at peer-reviewed AcademyHealth national meeting) is consistent with expert and other recommendations that to identify all ED visits, one also needs to include hospitalizations for asthma as potential indicators of an otherwise unrecognized ED visit, which we have done and incorporated into the specifications.

As the constructs of this measure are defined via the expert panel process and this is an innovative approach to measuring undesirable asthma outcomes, there is no gold standard or statistical analysis. As an outcome measure, no association with process needs to be tested, although the NHLBI guideline discusses ED visits and hospitalizations as undesirable and potentially preventable outcomes. We used Median Scores from the panel ratings of at least 7 to identify desirable constructs for the measure.

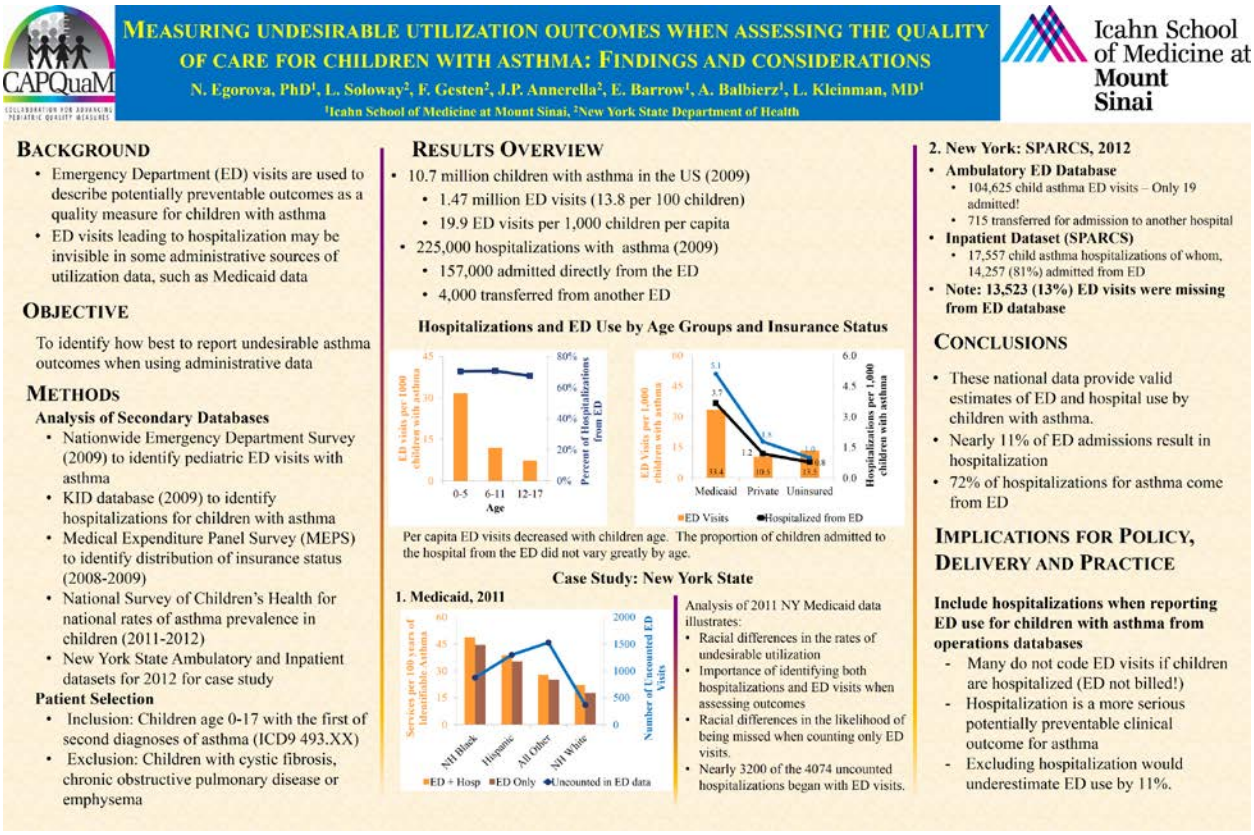
Some interesting exemplar ratings are shown below, with some key findings bolded:

Scenario	**MED**
In general, this measure is intended to describe care for children who have asthma and identifiable since before the ED visit.	9
Asthma is established by a single prior hospital admission with asthma as the primary discharge diagnosis	9
A single admission is not sufficient to establish the presence of asthma.	1
In children after their 5th birthday, Asthma is established by a single prior ED visit with asthma as the primary discharge diagnosis	8
In children after their 5th birthday, Asthma is established by a single prior ED visit with asthma as the secondary discharge diagnosis	7
In children after their 5th birthday, Asthma is established by a single prior ED visit with asthma as any discharge diagnosis	6
In children after their 5th birthday, Asthma is established by 2 or more outpatient visits with asthma as a diagnosis.	9
In children after their 5th birthday, Asthma is not established until 4 or more outpatient visits with asthma as a diagnosis	2
Asthma related medication use helps to establish the presence of asthma.	**8**
Prescription for leukotriene inhibitors are typically asthma related..	**5**
Prescriptions for long acting beta 2 agonists are typically asthma related.	9
Prescriptions for inhaled steroids are typically asthma related.	8
Oral steroid bursts are typically asthma related.	5
In order to establish a diagnosis of asthma, a child should experience a total of at least 2 asthma related events such as outpatient visits for asthma and or asthma related prescriptions, one of which must be an outpatient visit	8
Filled prescriptions should not be considered when establishing the presence of asthma	1
Children with a diagnosis of COPD with chronic aspiration should be excluded from this measure.	**9**
Children with a diagnosis of COPD should be excluded from this measure.	**9**
Children with a diagnosis of cystic fibrosis should be excluded from this measure.	**9**
Children with a diagnosis of emphysema and chronic aspiration should be excluded.	**9**
Children with a diagnosis of emphysema should be excluded	**9**
The time frame for establishing a diagnosis of asthma extends before the reporting year.	**9**

This measure should include children over 2	8
The upper age limit for this measure should be children until their 20th birthday	4
The upper age limit for this measure should be children until their 21st birthday	**9**
For reporting purposes, adolescents 19-21 should be grouped with adolescents under 18.	5
For the purposes of this measure, only ED visits with asthma as the primary diagnosis are eligible for inclusion.	3
For the purposes of this measure, only ED visits with asthma as the primary or secondary diagnosis are eligible for inclusion.	**8**
For the purposes of this measure, all ED visits with asthma as a diagnosis are eligible for inclusion.	5
For the purposes of this measure, a treatment for asthma must be provided or prescribed in order for the ED visit to be eligible for inclusion.	3
In children prior to their 5th birthday, Asthma is established by 3 or more outpatient visits with asthma as a diagnosis	**9**
In children prior to their 5th birthday, Asthma is not established until 4 or more outpatient visits with asthma as a diagnosis	3

Our approach to identifiable asthma was validated by comparing the prevalence of identifiable asthma to the number of children with NY asthma claims, and to the prevalence estimate expected via analysis of the National Survey of Children's Health and to the prevalence of children with preventable asthma as defined by the NCQA's asthma measures. We sought to have a measure that would be much more inclusive than the persistent asthma criteria but still filtered with a threshold requirement. Indeed our findings supported this with more than 25% of all children with asthma claims eliminated by our definition, a denominator that was about 50% of the estimated survey-reported lifetime incidence of asthma and 2.8 fold the number of children included than the NCQA criteria. We note that the NSCH survey prevalence exceeded the single claim approach.

The following poster was presented at AcademyHealth and provides additional data that illustrates the importance of using both ED visits and hospitalizations to identify the rate of ED visits. In NY State using ED visits alone would miss about 13% of ED visits, nationally about 11%. The inclusion of hospitalizations will overestimate the number of ED visits by between 4 and 5 percent. As many of these hospitalizations are for acute exacerbations, the construct of undesirable utilization outcomes would include them, so that while the estimate is likely to be a bit high for ED visits, it is a fair estimate of asthma outcomes. Our approach to avoid de-duplication and double counting of an ED visit and its associated admission as two numerator events is specified (admission on same or next day in the same institution) to favor sensitivity over specificity and hence will slightly underestimate numerator events, thus reducing the overestimation from the inclusion of the hospitalization.



2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Please see the section above. The face validity of our expert panel, the test-test reliability, the critical importance of having a standard, reliable, and valid approach to measuring the rate of asthma ED visits all support this measure.

We interpret our measure to be a valid estimate of the rate of ED visits and an even better estimate of undesirable outcomes from asthma.

Our interpretation is that administrative data are reliable for identifying asthma, and that year to year test retest reliability seems to indicate similar patterns of performance when identifying ED visits for asthma, reinforcing the reliability of our operational definitions for identifying eligible children. Our specification provide a sensitive and face valid approach to identifying an unbiased sample of children with ED visits (ensuring we don't bias the results towards the inappropriate by missing those with hospitalization).

Most databases contain consistent elements, are available in a timely manner, provide information about large numbers of individuals, and are relatively inexpensive to obtain and use. Validity of many databases has been established, and their strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented (30). Administrative data are supported, if not encouraged by federal agencies, such as NIH, AHRQ, HCFA, and the VA. The Centers for Medicare & Medicaid Services has made clear to the participating AHRQ-CMS CHIPRA Centers of Excellence funded to develop measures in the Pediatric Quality Measures Program that it places a premium on feasibility when assessing those measures that it will most highly recommend to states to complete. The sources of data for the existing measure and other similar measures are typically based upon administrative data as well, providing consensual validation for using administrative data as the primary data source.

2b3. EXCLUSIONS ANALYSIS

NA □ no exclusions — *skip to section 2b4*

Exclusions are clinical and specifically guided by the explicit criteria developed by the expert panel.

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Denominator Exclusions: Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 Code: 496), Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03, 277.09), or Emphysema diagnosis (ICD-9 code 492xx). Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month, as well as the index reporting month itself.

There are no numerator exclusions.

Exclusions were only included if they were endorsed by the expert panel. In studying the denominator we found that a very few percent of potentially eligible children (<=2.5%) were excluded by clinical diagnoses. The use of three months of continuous enrollment was recommended by our multi-stakeholder consortium and avoids the exclusion of more than 20% of otherwise eligible children from the population with identifiable asthma compared to a 12 month requirement.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In order to develop a sample of approximately 125,000 children with asthma in our initial field test (that required a 12 month continuous enrollment criterion), we excluded 2121 with COPD, 650 with cystic Fibrosis and 482 with emphysema (those children were not mutually exclusive, in other words, children may have been excluded for more than one reason so the total number of exclusions was at least 212 and less than the sum of the three diagnoses (between 1.6% and 2.5% of otherwise eligible children).

Had we used a 12 month continuous enrollment criterion, we would have excluded more than 20% of otherwise eligible children.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions are clinical and represent construct validity rather than statistical considerations.

The exclusions are purposeful and not statistical, and are based upon the findings of the expert panel. Noise is likely to be reduced by the exclusion of key diagnoses. Longer continuous enrollment requirements would harm validity since large number of children with real symptoms who are established and being managed for asthma would have been excluded.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** 2 **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Specifications for this measure requires stratification by age group and race/ethnicity. Several additional stratifications are optional but may be required by the accountability entity. These variables include rurality/urbanicity and county level of poverty.

Within age group, we specify a number of stratifications as we have done for all of our CAPQuaM PQMP measure. Absent clear biological evidence that ED visits should be more likely in any of the sub categories we have chosen not to adjust but to report both topline and stratified results.

We used stratification to allow for a granular understanding of performance. However the rate in the population assessed is the rate. Biological data do not support risk adjustment to control for patient characteristics on the variables of interest. The Pediatric Quality Measures Program which funded development of this measure requests that measures be specified to be able to identify disparities and differences by a variety of characteristics and this measure does that.

The NIH NHLBI NAEPP guideline notes that goals of care and definition of successful management are the same regardless of baseline presentation. Hence clinical risk adjustment is not appropriate.

As indicated by the NHLBI guideline (http://www.nhlbi.nih.gov/files/docs/guidelines/asthgdln.pdf page 38)

> ***"An important point linking asthma severity, control, and responsiveness is that the goals are identical for all levels of baseline asthma severity.*** *A patient who has severe persistent asthma compared to a patient who has mild persistent asthma, or a patient who is less responsive to therapy may require more intensive intervention to achieve well-controlled asthma; however, the goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by therapeutic intervention."*

For reasons other than controlling for case mix, we specify this measure to be stratified by age group and race/ethnicity as well as providing a top line analysis. Without such stratifications, racial and ethnic disparities (which have been found to be prevalent in children with asthma) might go unnoticed. The CHIPRA legislation that funded the development of these measures asks for the capacity to identify such disparities to be included in the measure specifications.

Specifications for further stratifications, such as by rurality/urbanicity and by county level of poverty are provided, in the event such stratification is required by the accountability entity or desired by the reporting entity.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk

(*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) Consistent with the Disparities Working Group of the Pediatric Quality Measurement Program, CAPQuaM has chosen an approach to not risk adjust for outcomes as being most appropriate to measuring actual performance. Nonetheless, we honor the parameters in the legislation funding the PQMP and also recognize the interest of various stakeholders in comparing like-to-like: hence we have specified key stratifications for analysis and presentation. The accountability entity has the option to request the granularity of stratification that suits its needs beyond age strata and race/ethnicity.

The conceptual model is that of CAPQuaM that includes that in pediatrics age is a key predictor and stratification is valuable. We were asked by AHRQ and CMS to include other constructs and we have manifest them as specified, such as race/ethnicity, poverty level in the caregivers county of residence, rurality/urbanicity on the caregiver's county of residence, insurance type and plan type, when variable. We have not added a stratum for children with special health care needs since asthmatics going to the emergency room are highly likely to belong in this category.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

N/A

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

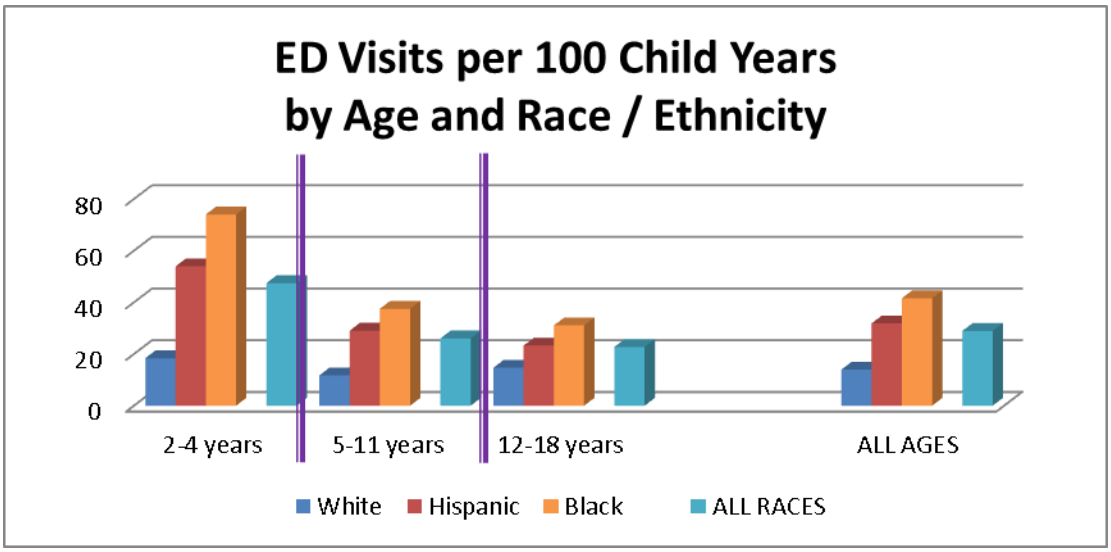
2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
2b4.9. Results of Risk Stratification Analysis:

Data presented in chart and figure show asthma outcomes stratified by age and race/ethnicity. Additional analyses showed meaningful stratifications by time of year, county level of poverty, and rurality/urbanicity.

Stratification by Age and Race/Ethnicity		
Age Group	**Race / Ethnicity**	**Rate per 100 Child-Years**
2-4 years	**Non Hispanic:**	
	White	18.40
	Black	74.03

	Asian	19.29
	Other	51.11
	Hispanic	53.93
2-4 years	**TOTAL**	**47.44**
5-11 years	**Non Hispanic:**	
	White	11.74
	Black	37.51
	Asian	10.18
	Other	26.15
	Hispanic	15.19
5-11 years	**TOTAL**	**26.03**
12-18 years	**Non Hispanic:**	26.75
	White	14.68
	Black	31.07
	Asian	7.42
	Other	24.56
	Hispanic	32.00
12-18 years	**TOTAL**	**22.17**
19-21 years	**Non Hispanic:**	36.27
	White	25.86
	Black	48.10
	Asian	8.70
	Other	28.00
	Hispanic	32.82
19-21 years	**TOTAL**	**32.60**
OVERALL	**TOTAL**	**34.10**



All of these are statistically significant by chi square analysis, p well below 0.05.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

We acknowledge the association of the stratification variables with the rate of asthma ED visits but have not found evidence justifying such differences as either acceptable or un-modifiable by health care. Indeed there is evidence that primary care, adherence to guidelines, and other healthcare interventions can reduce or eliminated the impact of these factors. Federal guidelines quoted above support this perspective.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Contingency table analysis with chi-square and using t-statistics were coherent and each illustrated the presence of statistical differences.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Differences between major subgroups were statistically significant, including race/ethnicity, age group, level of poverty in the county, and level of urbanicity (urban, suburban, rural). All one way and two way (within age stratum) chi square analyses and t test analyses were $p<0.05$. As noted below this is a stricter test than had the measure been assessed across different entities. We have requested additional analysis that would compare vendors within the NY State Medicaid system but we are not sure if the NY State Department of health will be able to complete these analyses in time for the Committee meeting (the data are owned and managed by New York State, who recently changed the organization of their raw data base tables, delaying their capacity to respond to our request). We note that these analyses will not produce the beta binomial signal to noise ratio frequently submitted with NQF measures as the measure itself is not bounded by 0 and 1 as required by the beta binomial model.

We further note that the measure was sensitive enough to demonstrate face validity with statistically significant differences from month to month and season to season as expected for this outcome. Within the NY State Medicaid data, differences were also found by eligibility category, which in this case can serve as proxy for health plan. Chi Square of the rate difference (7.7 ED visits per 100 children) between those qualifying for cash assistance versus those qualifying because of SSI was 32.07 with one degree of freedom meaningfully exceeds the critical value of 10.828 for $p<.001$. This demonstrates excellent capacity to distinguish between health plans. County by county results also varied significantly and were aggregated by urban influence code and level of poverty and continued to show capacity to show meaningful differences.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The measures are sensitive enough to detect meaningful differences as observed within a population (as with the seasonal and month to month variations described above). Since the sum of squares across populations is expected to be greater in distinct populations, we expect the measure to perform very well when comparing across populations as well. Since the effective sample size of within population comparisons (such as we have conducted) is diminished by an (unmeasured) intraclass correlation coefficient, we would expect greater power for equal sample size to detect differences between entities than we had in our testing of various subpopulations within a single state. This supports the same conclusion. The signal to noise ratio is very strong for this measure.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

We use administrative claims to establish eligibility. No missing data analysis performed as we were using standard data sources that are contractually obligated to be provided to NYS Medicaid. Our analyses found that the absence of pharmacy data would reduce only slightly the number of children identified as having identifiable asthma. This finding became apparent during alpha testing of our specifications and was incorporated into our specifications as a permissive allowance when pharmacy data were not available. We are unable to quantify the reduction in denominator at present because NY State (the owner of the data used for these analyses) recently changed their data tables and have not yet had the capacity to complete CAPQuaM-requested analyses.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Addressed elsewhere regarding data sources and definition of identifiable asthma, requirements for 3 months of continuous enrollment. The use of a composite requirement to establish eligibility reduces that likelihood of systematic error or dependence upon any specific data field. The use of complementary sources of identifying visits (CPT codes and revenue codes) accomplishes a similar goal.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Generally N/A. Systems unable to integrate pharmacy data into the eligibility analysis would have a minimally higher risk population than those with pharmacy claims. The specifics of the definitions and the limited impact of pharmacy claims on eligibility combine to make the expected impact of this on the rate of ED visits to almost

zero. They are included in the identification of denominator because our expert panel directed us to do so. More importantly, as cited above, the NHLBI guideline tells us that outcomes should not be adjusted for baseline risk, so this does not truly disadvantage a reporting entity according to the guideline.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other

If other: In rare instances, race/ethnicity and zip code may need to be abstracted from the chart.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Medicaid systems typically include race and ethnicity and zip code as defined electronic fields.

Emergency department visits that result in hospitalization are often not coded as ED visits in administrative data. The most valid estimate of ED visit rate requires use of both ED and hospitalizations as numerator events.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

None at present

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Public Health/Disease Surveillance Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization) Not in use	

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

The measure is not currently in use.

The topic of ED asthma use was assigned to our measure development project in the Pediatric Quality Measures Program by CMS, by far the largest single third party payer for medical care for children in the US, and by AHRQ. Major federal policy makers have indicated to us that these measures are a priority. This measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

We have begun discourse with the CDC regarding use of this measure.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

As a part of our work with PQMP, we are working on specific plans for dissemination and use. Our plan for implementation includes submitting our application for measurement endorsement from the National Quality Forum. We are having conversations with NY State Medicaid (who was one of our partners in development) regarding the application and use of this measure. No time frames

have been established.

Meeting the expected timeframes of NQF, CAPQuaM intends for the measure to be used for an accountability application within 3 years of initial endorsement and public reporting within six years of initial endorsement.

At this point in time, the submitted measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

We have begun discourse with the CDC regarding use of this measure.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Not Applicable.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Rates of emergency department visit use for children managed for identifiable asthma is an important outcome measure with intrinsic value that helps ensure high-quality, efficient healthcare for individuals and populations. Not all ED visits for asthma are necessary, some cases require a different level of care for the clinical circumstance. Also, a significant proportion of visits potentially could have been prevented with better prior management. A variety of stakeholders benefit from this measure:

- Plans could provide clinicians this data to use to accurately identify patients who benefit from enhanced asthma care.

- Health systems can use this data to distinguish patients who have identifiable asthma, their demographics, and the care they receive and the associated costs. This information allows practices, groups and facilities to evaluate and compare treatment plans between practice sites, medical and other professional groups and between integrated or other delivery networks. This evidence-based evaluation promotes the adoption of more effective and efficient health systems.

- States and healthcare agencies can also use these measures to compare larger systems to test and evaluate treatment options, payment models (e.g. managed care, primary care case management), quality of health plans, costs and health outcomes. Findings can be stratified by state or regionally (e.g. urban, rural, health shortage regions) to understand policy, demographic and culture effects.

- The data also allows clinical, public health and epidemiology researchers to understand the type, level and cost of care patients are receiving related to their health outcomes, giving opportunity to compare between health plans, payment models and treatment options. It also gives a deeper understanding of how individual (e.g. sex, age, gender, race, social economic status, poverty) and community determinants (e.g. work environment, community benefits) affect the rate of ED visits over time. Furthermore, these measures gives researchers the tools to identify and reach out to patients and their families to understand their health culture and practices to ensure that health services offered will most likely be utilized.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative**

unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
There has not been any evidence of unintended negative consequences to individual or populations. There are no anticipated unintended consequences if measuring at the level of comparing states, geographic regions, payment models, or health plans. Comparing individual health care professionals is not recommended as care is provided across practices may be necessary. Also, it is not appropriate for a single hospital comparison because it is measuring the system performance not the hospital performance. Lastly, although the measure can be used to compare practice sites, medical or other professional groups or integrated or other delivery networks, the measures are only recommended for large practices or integrated delivery systems that own their own risk and manage inpatient and outpatient care or that have access to all payer data sources

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Our definition of identifiable asthma is more inclusive than, for example, NCQA's persistent asthma construct. We use similar medication definitions as NCQA, except we exclude leukotriene inhibitors from asthma-related medications because our expert panel felt that these medications were used frequently for allergy patients and judged that the small gain in sensitivity of identifying children (considering all criteria) would be less than the loss in sensitivity and likelihood to include non-asthmatic children with allergies. Our specifications have been validated by an expert panel in the context of a peer reviewed process commissioned by AHRQ and CMS to advance the field and science of pediatric quality measurement beyond the state represented in pre-existing measures. The specification of a person-time denominator allows for the measure to have a shorter requirement for continuous enrollment than other measures with less risk of bias than previous measures.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Asthma_rate_Appendix__final-635857041553311257.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University Hospitals Cleveland Medical Center
Co.2 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 617-699-3357-
Co.3 **Measure Developer if different from Measure Steward:** Collaboration for Pediatric Quality Measures (CAPQuaM)
Co.4 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 617-699-3357-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Role: Expert Panelists
Elizabeth Allen, M.D Ohio State University
Chitra Dinakar, M.D. Department of Pediatrics University of Missouri-Kansas City
Stephen Teach, M.D., M.P.H. Children's National Medical Center
Charles Macias M.D., M.P.H. Pediatrics, Baylor Texas Children's Hospital
Michael Cabana M.D., M.P.H. University of California, San Francisco

Barbara Yawn M.D., M.S. Olmstead Medical Center, University of Minnesota
Joan Connell M.D. University of North Dakota
Delaney Gracy M.D., M.P.H. The Children's Health Fund and Montefiore Children's Hospital
Sharlene Miner M.D. Emergency Medicine Inter Mountain Medical Center

ROLE: Steering Committee and Investigator
Wilson Pace, MD American Academy of Family Physicians – DARTNET Institute - University of Colorado
Lynn Olson, PhD American Academy of Pediatrics
Christina Bethell, PhD, MBA, MPH Child and Adolescent Health Measurement Initiative, Johns Hopkins University (Previous OHSU)
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Lawrence Kleinman, MD, MPH Icahn School of Medicine at Mount Sinai
Rebecca Anderson Mount Sinai Medical Center
Ira Nash, MD Northshore - Long Island Jewish Medical Center (previous Mount Sinai School of Medicine)
Eyal Shemesh, MD Icahn School of Medicine at Mount Sinai
Mary Barton, MD National Committee on Quality Assurance
Charles Homer, MD, MPH US Department of HHS (previous National Institute for Child Health Quality)
Marla Clayman, PhD American Institutes for Research (previous Northwestern University)
Foster Gesten, MD New York State Dept. of Health, Office of Health Insurance Programs
Jerod M. Loeb, PhD The Joint Commission
Robert Rehm National Committee for Quality Assurance
Steve Kairys, MD American Academy of Pediatrics/QuIIN
Erin DuPree, MD The Joint Commission (previous: Icahn School of Medicine at Mount Sinai)
David Baker The Joint Commission (Previous Northwestern)
Beverley Johnson, BSN* Institute for Patient- and Family-Centered Care

Doris Peter Consumers Union

ROLE: Senior Advisory Board Member and Investigator
Shoshanna Sofaer, DrPH American Institutes for Research (previous CUNY Baruch)
Lynne Richardson, MD Icahn School of Medicine at Mount Sinai
Ian Holzman, MD Icahn School of Medicine at Mount Sinai
Marilyn Kacica, MD New York State Dept. of Health, Division of Family Health

ROLE: Senior Advisory Board Member
Marc Lashley, MD Allied Pediatrics
Gary Mirkin, MD Allied Pediatrics
Barbara Kupferman, RN, CMCN, LNC AmeriChoice by United Healthcare
L. Gregory Pawlson, MD, MPH SE Healthcare QUALITY Consultants (previous Blue Cross Blue Shield)
John Santa, MD* (previous Consumers Union)
John Clarke, MD ECRI, PA Patient Safety Authority
Scott Breidbart, MD Empire Blue Cross Blue Shield/ Anthem
Jeff Terry, MBA GE Healthcare
Robert St. Peter, MD, MPH Kansas Health Institute
Ruth Stein, MD Montefiore Children's Hospital
Arthur Aufses, MD Icahn School of Medicine at Mount Sinai
Eric Rose, MD Icahn School of Medicine at Mount Sinai
Wendy Brennan, MS National Alliance on Mental Illness of New York City (NAMI - NYC Metro)
Laurel Pickering, MPH Northeast Business Group on Health
Tony Hope, BM BCh, PhD Oxford University, The Ethox Center
Martin Hatlie, JD Partnership 4 Pt. Safety, Consumers Advancing Patient Safety
Paul Wise, MD Stanford University
Mark Chassin, MD, MPH, MPP The Joint Commission
Ira Nash, MD Icahn School of Medicine at Mount Sinai

ROLE: Investigators/Key Staff
Rusty McLouth, MS AAFP
Kasey McCracken, MPH CAHMI
Elise Barrow, MPH Icahn School of Medicine at Mount Sinai
Kasey Coyne, BA Icahn School of Medicine at Mount Sinai
Natalia Egorova, PhD Icahn School of Medicine at Mount Sinai
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Melissa Manice, MPH Icahn School of Medicine at Mount Sinai
Joshua Oppenheimer, BA Icahn School of Medicine at Mount Sinai
Eve Oster, MD Pediatric Partners of Gwinnett (previous Icahn School of Medicine at Mount Sinai)
Daniel O'Sullivan, BS University of Virginia (previous Icahn School of Medicine at Mount Sinai)
Michael Parides, PhD Icahn School of Medicine at Mount Sinai
Barbara Rabin, MHA Icahn School of Medicine at Mount Sinai
Carolyn Rosen, MD Icahn School of Medicine at Mount Sinai
Melissa Saperstein, MSW Icahn School of Medicine at Mount Sinai
Eyal Shemesh, MD Icahn School of Medicine at Mount Sinai
Andrew Ting, MD Icahn School of Medicine at Mount Sinai
Adam Vella, MD Icahn School of Medicine at Mount Sinai
Virginia Walther, MSW Icahn School of Medicine at Mount Sinai
Jennifer Zeitlin, SD Icahn School of Medicine at Mount Sinai, French Institute of Health and Medical Research (INSERM)
Marianne McPherson, PhD, MS NICHQ
Joseph Anarella, MPH NYS DOH, OHIP

Lee Sanders, MD, MPH	Stanford University
Kasey Coyne	Northwestern
Victoria Wagner	New York State Health Department Office Quality Patient Safety
Laura Soloway	New York State Health Department Office Quality Patient Safety
Kevin Malloy	New York State Health Department Office Quality Patient Safety
Judy Stribling	Icahn School of Medicine at Mount Sinai
Linda Radecki	AAP
Keri Thiessen	AAP – QuIIN
Laura Morris	New York State Department of Health
Olivia Lindly	OHSU
Kasey McCracken	OHSU
Sandeep Sharma	Icahn School of Medicine at Mount Sinai
Ann Nevar	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Suzanne Lo	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Amy Balbierz	Icahn School of Medicine at Mount Sinai
Samantha Raymond	Icahn School of Medicine at Mount Sinai
Allisyn Vachon	Icahn School of Medicine at Mount Sinai
Shannon Weber	Icahn School of Medicine at Mount Sinai
ROLE: Contractors	
Mary Fialkowski, MBA	The Joint Commission
Linda Hanold, MHSA	The Joint Commission
Elvira Ryan, MBA, BSN, RN	The Joint Commission
Sharon Sprenger, RHIA, CPHQ, MPA	The Joint Commission
Ann Watt, MBA, RHIA	The Joint Commission
Amy Slott	The Joint Commission
Denise Krusenoski	The Joint Commission
Elvira Ryan	The Joint Commission
Tasha Mearday	The Joint Commission

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:**

Ad.3 **Month and Year of most recent revision:**

Ad.4 **What is your frequency for review/update of this measure?**

Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2816
Measure Title: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma: A PQMP Measure
Measure Steward: University Hospitals Cleveland Medical Center
Brief Description of Measure: This measure estimates the proportion of emergency department (ED) visits that meet criteria for the ED being the appropriate level of care, among all ED visits for identifiable asthma in children and adolescents.
Developer Rationale: Asthma is one of the most common indications for emergency department (ED) visits by children. (1-3) AHRQ's Healthcare Cost and Utilization Project (HCUP) data from the Nationwide Emergency Department Sample (NEDS) found that in 2012, children between 1 and 17 years old had more than 1,895,000 ED visits for asthma with almost 10% resulting in hospitalization.
Evidence suggests that ED visits and hospitalizations in children with asthma vary systematically by how well-equipped that community is to provide primary care, and by the quality of primary care delivered. (4, 5) There is widespread literature illustrating that ED visits and hospitalizations are each undesirable utilization outcomes from poorly managed asthma. There is not a large literature that assesses whether or not pediatric ED visits were appropriate. (6 -10)

A body of literature has explored the value and feasibility of measuring the appropriateness of medical activities using data available in the medical record. (11-14) Early work in adults included assessment of hysterectomy, carotid endarterectomy and cardiac interventions. An independent research project brought the construct of appropriateness to children (15), while Kleinman and colleagues were the first to assess the appropriateness of specific pediatric procedures. (16, 17) A later study demonstrated the feasibility of medical record data for such an assessment. (18) DeAngelis pioneered studies of what constitutes a good reason to use the ED. (6) All of these studies used a definition of appropriateness that compared benefit to likely risk without specific consideration of costs. The need for more studies looking for overuse was recently reviewed. (19) RAND type Delphi panels are accepted around the world as a method for developing criteria to assess appropriateness. (20-22)

Research demonstrates that:

•ED visits are an important issue for child health insurers, including Medicaid, with clinical and financial consequences;
•An overcrowded primary care system contributes to ED use for non-emergent and even non-urgent conditions.
•Pediatric hospitalizations for asthma vary by primary care availability and quality
•ED visits are common for children with asthma, including those in Medicaid
•Assessment of appropriateness using information in the medical record is a well-established and validated method that has been successfully applied to children.

The literature suggests that a measure that assesses whether or not the ED is an appropriate level of care for a child with asthma at the time that they present has intrinsic value. Such a measure would:

•Characterize the process of care in a way that assesses whether a particular ED visit represents overuse
•Allow the outcomes of asthma care to be better characterized in a manner that describes performance and promotes targeted improvement. Inappropriate ED visits represent failures of primary care delivery, availability and/or access. Appropriate visits may represent a failure to control asthma. These have distinct and distinguishable meanings that contribute to the understanding of the quality of asthma care.
•Measuring the quality of asthma care requires assessment of multiple factors. This appropriateness measure helps plans, purchasers, and society to understand the implication of asthma ED visits as outcomes of asthma care. The implications herein is that understanding what is better or worse care requires looking at various factors and not simply a higher or lower appropriateness score. The understanding of this measure is enhanced by considering whether the rate of undesirable outcomes (ED visits and hospitalizations) is high or low and whether other measures of primary care availability and access or asthma quality suggest high levels of performance or not..

An abstract describing the proposed measure was peer-reviewed and subsequently presented to a national audience at AcademyHealth 2014 Annual Research Meeting in San Diego in the "Measuring the Safety, Quality, and Value" section. Feedback was positive regarding the methods, measures, ethics, and importance of this measure.

Research evidence supports the importance and need for our proposed measure that assesses whether the ED represents an appropriate level of care for children with asthma who are seen in the ED.

1.Kharbanda, A.B., et al., Variation in resource utilization across a national sample of pediatric emergency departments. J Pediatr, 2013. 163(1): p. 230-6.
2.Adams, J.G., Emergency department overuse: Perceptions and solutions. JAMA, 2013. 309(11): p. 1173-1174.
3.Institute, N.E.H., A Matter of Urgency: Reducing Emergency Department Overuse. Research Brief, 2010(March).
4.Perrin, J.M., et al., Variations in rates of hospitalization of children in three urban communities. N Engl J Med, 1989. 320(18): p. 1183-7.
5.Perrin, J.M., et al., Primary care involvement among hospitalized children. Arch Pediatr Adolesc Med, 1996. 150(5): p. 479-86.
6.DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p.61-5.
7.Berns, S.D., et al., Appropriate use of a pediatric emergency department: is the pediatrician called before the visit? Pediatr Emerg Care, 1994. 10(1): p. 13-7.
8.Rudowitz, R., A Look At CBO Projections For Medicaid and CHIP, in The Kaiser Commission on Medicaid and the Uninsured. 2014, The Henry J. Kaiser Family Foundation
9.Taubman, S.L., et al., Medicaid Increases Emergency-Department Use: Evidence from Oregon´s Health Insurance Experiment. Science, 2014. 343(6168): p. 263-268.
10.Smulowitz, P.B., et al., Increased Use of the Emergency Department After Health Care Reform in Massachusetts. Ann Emerg Med, 2014.
11.Brook, R.H., et al., A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care, 1986. 2(1): p. 53-63.
12.Park, R.E., et al., Physician ratings of appropriate indications for six medical and surgical procedures. Am J Public Health, 1986. 76(7): p. 766-72.
13.Fitch, K., et al., The RAND/UCLA Appropriateness Method User´s Manual. 2001 RAND.
14.Kosecoff, J., et al., The appropriateness of using a medical procedure. Is information in the medical record valid? Med Care, 1987. 25(3): p. 196-201.
15.Kemper, K.J., Medically inappropriate hospital use in a pediatric population. N Engl J Med, 1988. 318(16): p. 1033-7.
16.Kleinman, L.C., et al., The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. Jama, 1994. 271(16): p. 1250-5.

17.Kleinman, L.C., E.A. Boyd, and J.C. Heritage, Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians. Jama, 1997. 278(6): p. 497-501.
18.Keyhani, S., et al., Electronic health record components and the quality of care. Med Care, 2008. 46(12): p. 1267-72.
19.Keyhani, S. and A.L. Siu, The underuse of overuse research. Health Serv Res, 2008. 43(6): p. 1923-30.
20.Bernstein, S.J., et al., The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium. Jama, 1993. 269(18): p. 2398-402.
21.Taylor, A.J., et al., ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 Appropriate Use Criteria for Cardiac Computed Tomography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. J Cardiovasc Comput Tomogr, 2010. 4(6): p. 407.e1-33.
22.Basger, B.J., T.F. Chen, and R.J. Moles, Validation of prescribing appropriateness criteria for older Australians using the RAND/UCLA appropriateness method. BMJ Open, 2012. 2(5).

Numerator Statement: The numerator is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. Distinct numerators are reported for children ages 2-5, 6-11, 12-18, and optionally, 19 - 21.

Denominator Statement: The denominator represents a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (Appendix Table 1).

Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. An overall rate across strata is not reported.

Denominator Exclusions: ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema. Identifiable asthma is defined is section S.9.

At the discretion of the accountability entity, the denominator may be restricted to children 2-18.

These details incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
Level of Analysis: Health Plan, Integrated Delivery System, Population : Community, Population : County or City, Population : National, Population : Regional, Population : State

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report
1a. Evidence
1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Evidence Summary

The developer provides the following evidence for this process measure:

- The level of analysis is health plan, integrated delivery system, or population.
- The developer states the measure is "supported, but not defined by," a guideline from the National Heart, Lung, and Blood Institute (NHLBI) clinical practice guidelines for whether the Emergency Department (ED) represents an appropriate level of care for children with asthma who are seen in the ED:
    - *As a general rule, patients with well-controlled asthma should have:*
        - *Few, if any, asthma symptoms.*
        - *Few, if any, awakenings during the night caused by asthma symptoms.*
        - *No need to take time off from school or work due to asthma.*
        - *Few or no limits on full participation in physical activities.*
        - *No emergency department visits.*
        - *No hospital stays.*
        - *Few or no side effects from asthma medicines.*
- The developer does not provide information on the quality, quantity, and consistency of evidence in the guideline or whether it was graded.
- The developer also conducted its own review of peer-reviewed and grey literature from 1985-2014 and reviewed 91 articles in developing its appropriate use criteria in conjunction with its expert panel. Eight consensus appropriate use criteria are specified:
    1. Hospitalization directly from the ED
    2. Documented physical findings consistent with respiratory distress, including:
        a) Labored breathing with retractions and/or evidence of accessory muscle use
        b) Markedly decreased breath sounds
    3. O_2 saturation level less than 90 percent on percutaneous assessment;
    4. An arterial blood (ABG) gas obtained (or ordered);
    5. Consultation ordered and obtained with a pulmonologist asthma specialist, an order of an arterial blood gas (ABG), or a consult with a pulmonary or asthma specialist;
    6. Parent/caregiver referred to the ED after evaluation from the PCP or other office/clinic;
    7. Parent/caregiver report of administering two or more doses of inhaled rescue medications without meaningful clinical improvement;
    8. Parent/caregiver report that the child was in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan.

Exception to evidence

- The developer does not provide empiric evidence for each appropriate use criteria.
- The developer states it used the RAND/UCLA appropriateness method.
- NQF provides specific guidance on evaluating appropriate use measures, as follows (page 44):
    - "If there is no empiric evidence, skip Box 10 and go to Box 11. The Committee should agree that the AUC method is a systematic assessment of expert opinion that the

benefits of what is being measured outweigh the potential harms (Box 11). If the Committee agrees that it is acceptable (or beneficial) to hold providers accountable for the performance in the absence of empiric evidence (Box 12), then rate as "insufficient evidence with exception."

Guidance from the Evidence Algorithm: 1 → 3 → 7→ 10 → 11 (highest eligible rating of INSUFFICIENT WITH EXCEPTION rating)

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *Are there, OR could there be, performance measures of a related health outcome OR evidence-based intermediate clinical outcome or process? [Evidence Algorithm Box 10]*
- *Is there evidence of a systematic assessment of expert opinion (e.g., national/international consensus recommendation) that the benefits of what is being measured outweigh potential harms? [Evidence Algorithm Box 11]*
- *Does the Steering Committee agree that it is OK (or beneficial) to hold providers accountable for performance in the absence of empirical evidence of benefit to patients? [Evidence Algorithm Box 12]*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states, "asthma is one of the most common chronic diseases in children, affecting an estimated 7.1 million children in the United States. In 2011, 4.1 million children suffered from an asthma attack or episode. It is the second most common reason (after allergy) for children to be classified as having a special health care need, accounting for nearly 38.8% of such children."
- The developer reports the following results sufficiently identified statistically significant differences between groups at specified levels, e.g., age groups, among racial/ethnic groups, and within age group among racial/ethnic groups:
  - 181 of 335 (54.3%) ED visits were deemed appropriate for children 2 to 5 years
  - 209 of 447 (43.8%) ED visits were appropriate for children 6 to 11 years
  - 165 of 341 (48.4%) visits were appropriate for children 12 to 18 years

Disparities

- The developer states, "Pediatric asthma is more prevalent in minority populations. Lifetime prevalence rates of asthma in Hispanic and African American children are 12.4% and 15.8% respectively."
- Based on its chart audits, the developer reports performance on the measure varies by race/ethnicity and that a Chi-square analysis confirms the differences are statistically significant. For example, Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (47.8%), p=0.002. African American children "showed a trend" toward more questionable use compare to all other children (53.6% vs. 48.7%, p=0.10).
- The developer reports performance on the measure varies by insurance status and that a Chi-square analysis confirms the differences are statistically significant. The appropriate use rates were: Medicaid patients (46.3%); private (59%); uninsured patients (38.6%); other forms of insurance (military and worker's comp) (55.0%), p=0.005.

Question for the Committee:

- *Does the Committee believe there is a gap in care that warrants a national performance measure?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**The developer has identified this measure as a process measure. Much of the discussion in the methodology as well as the literature referenced could also consider this to be an outcomes measure. The evidence does support the measure as an identifier of the effectiveness of the process (primary care effective treatment).

**As noted, evidence is lacking. however face validity of measure seems strong enough to accept.

**While the document characterized this as a "process measure" I believe this appropriateness measure reflects an outcome. Clearly the measure developers describe extensively the relationship between ED use for asthma patients and access to ambulatory care and ambulatory care quality.

**Almost no direct evidence reported for the bundle or groups of processes.

1b. Performance Gap

Comments:

**Data reviewed by the measure developer indicated a gap in age, race/ethnicity and in rural/urban. The literature reviewed by the developer demonstrated opportunities for care improvement based on disparities in care.

**Yes, a gap appears to exist.

**In their initial analyses of NY Medicaid data, 44-54% of asthma ED visits were deemed appropriate by the measure suggesting a substantial opportunity for improvement. However the analysis was done in a single state and no benchmark data were available (the target rate is not 100%).

**Little debate that gaps exists, but assessing it will be a challenge and make induce documentation vs care gap conflation.

1c. High Priority (previously referred to as High Impact)

Comments:

**This is not a composite performance measure. Not applicable.

**na

**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

Specifications:

- The numerator of this measure is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level

of care. The developer requires reporting by age, as follows: children ages 2-5, 6-11, 12-18, and optionally, 19-21 years.

- The denominator of this measure represents a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (Appendix Table 1).
- The developer states that separate numerators and denominators also are required reporting for: children age 2-5, 6-11, 12-18, and, optionally, 19-21 years.
- The ICD-9 and ICD-10 codes have been included in the specification details.
- The calculation algorithm is stated in S.18.
- One data source is pharmacy claims, but the developer acknowledges that availability will vary.

Questions for the Committee :

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*
- *Is the potential variability in access to/inclusion of pharmacy a concern?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developer relied on pre-existing data element-level validity testing to identify children who are being managed for identifiable asthma (denominator).
- The developer relied on empirical testing at the data element level (chart abstraction compare to an authoritative source) to assess the numerator reliability.
- Per NQF guidance, separate reliability testing is not required if data element-level validity testing is performed.

Results of reliability testing

- Not applicable; see discussion on validity testing at the data element level

Guidance from the Reliability Algorithm Not Applicable

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

- The numerator of this measure is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. The developer requires reporting by age, as follows: children ages 2-5, 6-11, 12-18, and optionally, 19-21 years.
- The denominator of this measure represents a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (Appendix Table 1). Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. The developer states an overall rate across strata is not specified for reporting.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer reports data element level validity testing. The developer relies, as is permitted by NQF guidance, on other sources for denominator data element validity and conducted empirical testing on the numerator data elements.
- The developer also cites face validity, but did not specifically assess face validity at the computed measure score level, as required by NQF for face validity testing.

Denominator

- The developer relies on literature to support its conclusion of the validity of administrative data elements to identify children who are being managed with identifiable asthma. Per NQF policy:
  - Validity testing at the data element level obviates the need for reliability testing at the data element level provided the critical data elements of the measure as specified are tested.
  - Prior evidence of validity of data elements can be used, including published data, provided it includes the same data elements; uses the same data type; and is conducted on an appropriate sample (i.e., representative, adequate numbers, etc.)
  - The developer attests that the data elements match those assessed in the literature.
- The developer also cites two NQF-endorsed measures from NCQA (*NQF 1799: Medication*

Management for People With Asthma and *NQF 1800: Asthma Medication Ratio*, as well as *NQF 0036: Use of Appropriate Medications for People with Asthma*, which is no longer being maintained) as evidence of data-element level validity.
    - The developer acknowledges #1799 and #1800 are not directly applicable because they were tested at the score level. The developer notes, however, "the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to [reliably] produce valid scores."
    - The developer states there is "nearly complete overlap of the denominator codes for this measure and there is overlap of the denominator elements." The developer states that where codes differ, they were specific decisions by its expert panel.
- The developer used NY State Medicaid Managed Care claims data for its analyses.
- The developer conducted empirical analyses on identifying ED visits.

Numerator

- For its validity testing at the data element level, the developer conducted 30 comparisons by two abstractors per construct at a single site against the conclusions of a physician reviewer (the authoritative source). The developer reports:
    - "Six constructs were for specific reasons for appropriateness (like markedly reduced breath sounds or evidence of respiratory distress."
    - The seventh construct was the visit-level assessment of appropriateness.
    - The developer conducted testing at the beginning of data collection and again at the conclusion of data collection.

Validity testing results:

Denominator

- As noted earlier, the developer refers to NCQA's data for three measures—#0036, #1799, and #1800—because these measures also rely on "nearly complete overlap" in the codes necessary to identify the target population of children with identifiable asthma.
    - The developer cites these measures as evidence that administrative data have the capacity to appropriately identify patients with asthma because the measure scores distinguish signal from noise. That is, were these measures inadequately able to do so for their denominator populations, the empirical reliability testing would not have yielded high reliability scores.
- The developer provides information about various articles related to the use of administrative data for performance measurement.
- The developer conducted empirical analyses on identifying ED visits and concluded that hospitalizations for asthma were potential indicators of an otherwise unrecognized ED visit, and excluding them would underestimate the denominator by 11%. Hospitalizations for asthma were incorporated in the specifications, but "final inclusion requires evidence of an ED visit."

Numerator

- Each construct is not individually reported, but the developer reports results from the initial assessment were:
    - Prior to training: PPV ranged from 0.78-1, NPV from 0.71-1, Se from 0.60-1.0 ,and kappas were 0.44, 0.6-0.60. 0.67,0.79, and 1.0.
    - Aggregating the 180 comparisons, PPV=0.89, NPV=0.91, Se=0.76, Sp=0.96, and

kappa=0.757, indicating very good to excellent agreement at the data element level.

- The developer reports the following testing results for the seventh construct of assessing appropriateness for each visit, for which there were 30 comparisons:

PPV	0.95
NPV	0.80
Sp =	0.888889
Se=	0.904762
Kappa	0.769231

- The developer states the assessment conducted at the end of data collection yielded comparable results. The appropriateness of visit results were:

PPV	0.882353
NPV	1
sp	0.866667
se	1
Kappa	0.866667

- The developer concludes the measure is be a valid estimate of the rate of ED visits and an even better estimate of undesirable outcomes from asthma.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following:

- There are no numerator exclusions.
- Denominator exclusions include: Children with concurrent or pre-existing Chronic Obstructive Pulmonary Disease (COPD) diagnosis, cystic fibrosis diagnosis, or emphysema diagnosis.
- The developer reports (<=2.5%) potentially eligible children were excluded by clinical diagnoses.
- The developer reports that exclusions are clinical and represent construct validity rather than statistical considerations.

Questions for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☐ **Statistical model** ☒ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

The developer provides the following information:

- "Specifications for this measure require stratification by age group."

- The developer found the measure varies by race/ethnicity, as follows, but states due to the lack of clear biological evidence, it has chosen not to adjust this process measure:
    - Appropriateness varied by age (Chi^2=8.2,p=.02), with younger (p=.01) and school aged (p=.01) children each being significantly different. Adolescents did not show a significantly different result when compared with all other ages.
    - Race/ethnicity: Hispanics at 44.1% appropriateness, non-Hispanic Blacks at 51.3%, Whites at 56.5% and all others at 72.2%. Chi square with 3 degrees of freedom was 15.4, with p=.0015. The appropriateness of ED visits for Hispanic children was less than for other children (p=.002).
    - Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (46.8%), p=.002. African American children showed a trend toward more questionable use compared to all other children (53.6% questionable vs 48.7%, p=.10).

Questions for the Committee:

- *If a justification for no risk adjustment is provided, is there any evidence that contradicts the developer's rationale and analysis?*
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on its analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer analyzed claims data from a single source, New York State Medicaid Managed Care data (including claims from all MCOs that are contracted for Medicaid care), so could not analyze differences among (for example) state Medicaid programs. The developer does not examine differences among MCOs within the data plan. The developer did examine meaningful differences in subpopulations within the state.
    - The developer performed contingency table analysis with chi square, generalized linear models and logistic regression analyses, each illustrated the presence of statistical difference among identifiable subgroups.
    - The developer states its analyses found meaningful differences by age groups and true signal in social determinants, consistent with the asthma literature, and did not incorrectly identify weak signal as meaningful.
- The developer posits that since the measure is sensitive enough to detect meaningful differences as observed within a population, the sum of squares across populations is expected to be greater in distinct populations and so the measure should perform well across populations.

Question for the Committee:

- *Does this measure identify meaningful differences about quality? At what level of analysis?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- The developer does not account for missing data. It cites literature that chart review is an accurate method of identifying the level of appropriateness of a clinical service. Failure to document is a "quality deficit" that the developer does not consider as missing data.
- Use of pharmacy data is on an "if available" basis to identify children with asthma for the denominator; the developer notes any results reported without should be marked as such. The developer reports use of pharmaceutical data expanded the pool by approximately 10,000 children (from 180,000 to 190,000—5.5%). The developer states it "found no evidence this was a threat to validity," but does not provide analyses that the 'pharmacy data cases' did not differ from the initial population. The developer does not have direct access to the data to provide additional analyses at this time.

Question for the Committee:

- *Is the variable use of pharmacy data a threat to validity?*

Guidance from NQF Validity Algorithm 1 → 2 (no demonstration of meaningful difference at Health Plan/Integrated System levels)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**The developer utilized data element validity testing. The developer also cites face validity, but did not specifically assess face validity at the computed measure score level, as required by NQF for face validity testing.
**2b1 seems adequate
**The specifications seem consistent with the evidence.
**The age grouping have no clear basis, and inclusion of (potentially) 19 years and above is a distractor, not an aid - and covered in overlap elsewhere.

2a2. Reliability Testing

Comments:

**No concerns with the validity testing. The developer conducted empirical analyses on identifying ED visits and concluded that hospitalizations for asthma were potential indicators of an otherwise unrecognized ED visit, and excluding them would underestimate the denominator by 11%. Hospitalizations for asthma were incorporated in the specifications, but "final inclusion requires evidence of an ED visit." The developer concludes the measure is be a valid estimate of the rate of ED visits and an even better estimate of undesirable outcomes from asthma.
**appears adequate. Agree that as a measure that the score represents a measure of quality, on a case by case basis there are likely to be appropriate reasons for ED visits.
**While the scope of testing was adequate, it was performance on HCUP data from a single state and may not reflect care in other settings.

2b2. Validity Testing

Comments:

**The exclusion criteria are appropriate for the measure and are identified by ICD codes. There is no risk-adjustment applied though the data are stratified by age for reporting purposes. The only meaningful differences would be by age and if stratification by race/ethnicity, urban/rural were to be conducted. Administrative data was utilized for this measure so missing data is not a concern.
**Case level metrics would be challenging yet aggregated scores would likely represent a better measure of quality. however this is a threat to validity.
**The absence of pharmacy administrative data may miss some patients that should be identified in the

denominator for asthma ED visits, but the proportion of patients missed should be low.
**Poor missing data approach - essentially ignored.
Pharmacy data variability is large threat.
Age group stratifications seem arbitrary; it is clear that a 4 year old and a 16 year old population differ in many ways, but these crude compartments are of ? value.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**The developer referenced NCQA measures as support for the reliability of the measure. The developer provides information about various articles related to the use of administrative data for performance measurement.
**na
**Reliability testing was done at the data element level. The testing was of adequate scope.
**Despite my pre-review concern, the developers make a solid case for reliability with their testing.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer reports some data elements are in defined fields in electronic sources.
- The developer reports there are no fees.
- The developer note that chart review is a reliable and accepted method of measuring appropriate use. No information is provided on the minimum number of charts that should be assessed.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**The data to calculate this measure come from administrative data (claims or encounters). It would be feasible to also consider moving this towards an e measure as the fields necessary should be included in an electronic health record.
**seems feasible
**While a substantial proportion of the data needed for this measure will be available in electronic format (claims or EHR data), some of the physical findings needed to identify appropriate ED care will not be found in structured fields of the EHR. Operationalization of this measure will require medical record review.
**No data about feasibility of all data being available...and given charting and recording variances, this is large threat to spurious observations.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The developer plans to assist in the implementation of this measure following NQF endorsement. The developer notes multiple stakeholders are interested in using the measure. It has been approved for inclusion in the National Quality Measures Clearinghouse.

Improvement results

- This measure is not in use and therefore, has no improvement results.

Potential harms

- The developer reports no unintended consequences were observed during testing.

Feedback: No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**The measure is not currently publicly reported. The measure results should lead to opportunities for interventions and activities that further the improvement of high quality, efficient primary care service delivery. No unintended consequences were noted by the developers.

**Used as a publicly reported measure, this could offer improvements in care.

**While the measure can be reliably calculated and will most assuredly show variations in "appropriateness" of ED visits for asthma, the lack of adjustment for sociodemographic factors will limit the usefulness of the measure for accountability purposes. Recent relevant studies show that the symptoms a patient presents with may not predict whether the "non-emergent" ED visit will be appropriate or not. The use of the area deprivation index may be useful to adjust data for sociodemographic factors.

Raven MC, Lowe RA, Maselli J, Hsia RY. Comparison of presenting complaint vs discharge diagnosis for identifying "nonemergency" emergency department visits. JAMA. 2013; 309:1145-53.

Kind AJ, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization: A Retrospective Cohort Study. Ann Intern Med. 2014;161:765-774.
The measure has not yet been used for accountability.
**Little assessment of potential harm - esp. of specific variable testing/reaction by provider.
Otherwise, no concerns. No current measure in this age, and potential for future program apparent.

Criterion 5: Related and Competing Measures
Related or competing measures • 2852: Optimal Asthma Control • 2794: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: CAPQuaM PQMP Asthma V: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in

understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

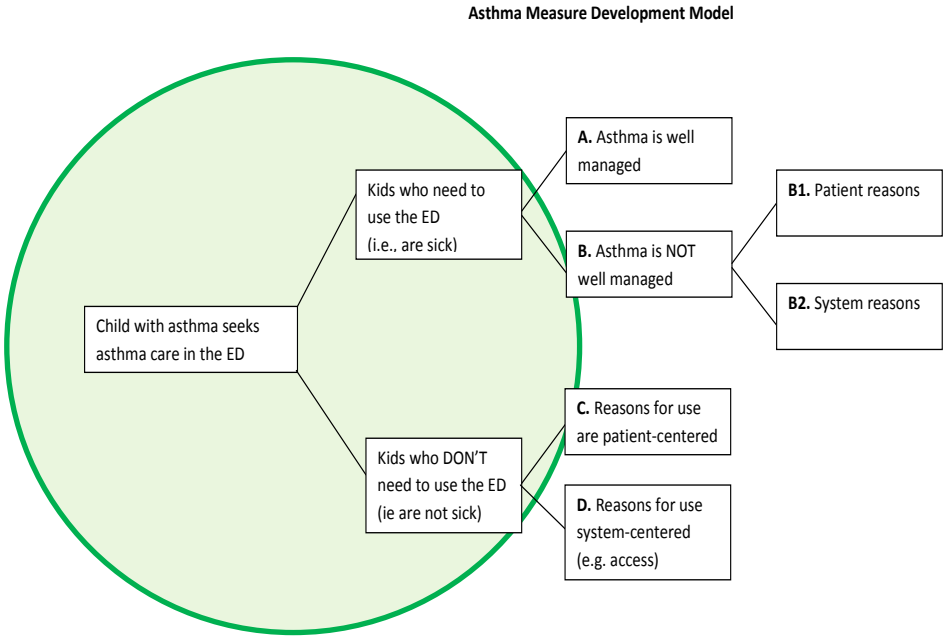
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



The green circle highlights the aspects of the conceptual model incorporated into this measure.

Underlying this model is a simple framework:

1. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who have acute breakthrough episodes requiring the ED.
2. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who come to the ED for asthma care better performed in the office setting.

3. Some children in the ED need to be there. Of those, some episodes were potentially preventable and others were not.
4. Our focus groups highlighted that some parents are comforted by the setting of the ED when they are caring for what they perceive as a vulnerable child. Parent perspectives do not adhere to system perspectives regarding a more strict hierarchy of what care belongs where.

Low levels of appropriateness suggest fewer breakthrough episodes of asthma and hence better quality of asthma care for those who receive it. If the rate of asthma ED visits is high and the rate of appropriateness is low this suggests both high quality care for those receive asthma care and insufficient access/availability of such care.

High levels of appropriateness suggest both efficient resource use of the emergency department and that ED visits are a proxy for clinical outcomes since many of the visits represent breakthrough asthma. High levels of appropriateness combined with a low rate of ED asthma use suggests both efficient use of resources and good asthma outcomes.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ **Other – *complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range:

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*) N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)? N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review. N/A

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

We conducted broad scoping reviews of our topic area: ED for asthma, overuse.

This measure is supported but not defined by the NIH NHLBI NAEPP Clinical guideline, that acknowledges ED visits as failures of control.

NHLBI Asthma Guideline 2006 www.nhlbi.nih.gov/guidelines/asthma (NAEPP Guideline)

Quick Reference: Asthma control focuses on two domains:

1)reducing impairment--- the frequency and intensity of symptoms… and

2) reducing risk – the likelihood of future asthma attacks… [later described as "prevent exacerbations]

NHLBI Guideline:

As a general rule, patients with well-controlled asthma should have:

- *Few, if any, asthma symptoms.*
- *Few, if any, awakenings during the night caused by asthma symptoms.*
- *No need to take time off from school or work due to asthma.*
- *Few or no limits on full participation in physical activities.*
- *No emergency department visits.*
- *No hospital stays.*
- *Few or no side effects from asthma medicines.*

We assert that without a measure of whether or not the reason the child is in the emergency room is sufficient to make it a clinically appropriate visit, it is impossible to interpret whether an ED visit represents a failure of clinical management and control, or a failure of the primary care or other aspects of the health care system to provide care at a more appropriate level of care.

We conducted a scoping review as follows:

We identified key constructs of asthma ED use measures for consideration. We created a table of these constructs in technical and lay language, and listed research questions for the review to answer. Our contractor (a national accrediting body experienced in measure development), prepared for us a literature review in 2 stages and we supplemented this with targeted reviews as needed to answer specific questions that arose during the measure development process.

The following construct table was used to guide the review and was the basis for the first round of review. Following the table, we include a list of questions for focused review that guided round 2 of the review, which resulted in a detailed summary of 91 articles from the peer-reviewed literature. In addition to this review, the CAPQuaM scientific team conducted an ad hoc series of reviews to answer specific questions such as the reliability of administrative data to identify asthma, and the value of expert panels and the RAND/UCLA appropriateness method. The CAPQuaM degree 360 method starts with a topic area and the measures emerge during the process, in this case necessitating the specified ad hoc reviews.

We searched peer reviewed and gray literature from 1985-2014 over the course of these reviews. Literature was summarized for our expert panel, which met in late 2013.

Overarching statement: Even when not specifically indicated, we are interested in how these constructs are impacted by such factors as race, ethnicity, socioeconomic status or its indicators, or the presence of other special health care needs.

Our metric is designed to capture axes related to two distinct conceptual frameworks:

1) Asthma is a model of chronic disease management. In other words, ED visits may arise from acute exacerbations indicating a flare up of disease, and/or suboptimal management of the chronic illness.
2) ED visits for asthma may reflect limitations of primary care beyond the provision of suboptimal treatment, such as insufficient education, limitations of access or availability, breakdowns of communication, or a variety of other factors.

We note that the internal quality of the ED visit to manage the asthma is not the target of this measure. However, communication between the emergency department and the primary care site may prove to be within the scope of this measure, pending the views of our experts and developers.

Construct I: Need to sufficiently specify population for measure

Concept	Implications (Lay Statement)	Lit Review Questions

(Descriptive) **The measure will need to adequately specify the population that we consider to be eligible for an ED with asthma measure.**	The development of measures regarding ED use for children with asthma requires us to understand the strengths and weaknesses for our measure of various approaches to identifying whether or not children have asthma. It further requires us to understand the impact of the availability of various sources of data (such as encounter data, pharmaceutical data, electronic medical record or chart review data) on these strengths and weaknesses. We are aware that the use of the term asthma is variable. We are not interested in diagnoses with the name asthma, but with an operational diagnosis that we will functionally treat as asthma, whether it has been called chronic wheezing, reactive airway disease, chronic infectious bronchitis, etc. We recognize that asthma and its presentation may change over the course of a child's life.	1. When asthma care is evaluated, how is the population of care recipients defined? How is asthma defined? What is the impact of including various types of data (dx 1 or more, drugs, etc) on the sensitivity and specificity of asthma identification? What are practical and valid approaches to identifying asthma? How do the answers to these questions differ between adults and children? 2. Are any groups persistently excluded from studies of asthma care (i.e., are children who have asthma and other comorbid conditions, such as a malignant disease, excluded?). What rationale is provided for the exclusion? 3. Are any non-asthma diagnoses considered to be indicators of asthma or potential asthma (e.g. bronchitis, bronchiolitis, wheezing, atopy) 4. For children up to age 21, how do issues of diagnosis, management, and follow-up differ by age and developmental stage? 5. At what point does literature suggest that reactive airway disease should be managed as asthma? a. What other conditions are managed as asthma? 6. What common current or preexisting comorbid conditions alter the management plan for asthma?

Construct II: Adequacy of management of asthma (as a chronic disease example)

Concept	Implications (Lay Statement)	Lit Review Questions
IIA. **↑Adequacy of asthma management:** **↓ED visits**	Since asthma is a chronic disease characterized by acute exacerbations, the extent to which asthma care is optimized through the use of appropriate medications, the control of the environment, and	1. What are the recommendations of the NHLBI guidelines? a. What does the literature suggest about the usefulness of NHLBI guidelines? b. Are there aspects that it has identified that appear to be missed? 2. What do we know about asthma management, how it's measured,

	the preparation of the parent/child dyad to adapt to changes in circumstances (e.g. viral respiratory infection or exposure to cold) should reduce the number of ED visits, irrespective of the number of primary care visits.	who provides it, patterns of care and how ED visits vary as a consequence? 3. Does identification of PCP improve outcomes of ED visit, including patterns of care, utilization? 4. What do we know about the content of an asthma plan and its relationship to a full program of chronic disease management, and its influence on ED utilization? 5. What evidence is there about the impact on outcomes such as ED use when the child or adolescent is involved in asthma self-management? For example, does it matter if: a. The child has a written asthma plan? b. The child understands their asthma plan? c. The child is given an opportunity to participate in managing care? 6. How is the role of the child in self-management measured? 7. How much are children able to recognize, communicate and act on their asthma? 8. What do we know about the impact of asthma services on asthma management? This includes: a. Treatment from an asthma specialistrand; b. Social worker; or c. Multidisciplinary personnel 9. To what extent is ED use by children with asthma stimulated by non-asthma related issues? a. How can we identify when that occurs? b. What is the evidence that providing other services will reduce the number of ED visits? 10. To what extent do children contribute to their management (including avoiding triggers, recognizing symptoms, medication adherence, etc.)? a. What is the impact and variance by age?

	11. What is the evidence regarding adequacy of various medication delivery systems for infants, toddlers, children and adolescents in acute and chronic settings? 12. Is there evidence of prior insult to the lungs such as sequelae of prematurity, etc. that create distinct subpopulations when considering this measure (at risk for ER visit)? 13. What aspects of the health services environment have been identified as contributing to outcomes of asthma management (e.g. school based health care)? 14. Does rate of ED utilization for non-respiratory diagnoses vary between asthmatics and non-asthmatics? 15. What is known about how often children with asthma use the ED over an extended period of time? Does it change over the life course of childhood? How does that vary by child characteristics, including race, SES, urban, suburban vs. rural, and age?

IIB. **↑PCP capacity/knowledge/skill:** **a. ↑Asthma management** **b. ↓Asthma exacerbations** **c. ↑Chronic disease management**	Broadly speaking, patient management of asthma is influenced by the capacity of the PCP practice. This includes the knowledge and skills possessed by the PCP, as well as office support to enhance access and availability of care. PCP includes the ability of the PC office to meet the cultural needs of the patient and their family.	1. What are the diversity of practices or services that may or may not impact ability or capacity of the PCP practice to manage asthma? 2. What do we know about the specific skills and processes that contribute to a primary care practice's capacity? 3. What patterns of visits or medication use or other indicators have been used as markers of well or poorly delivered primary care for asthma in children and/or adults? 4. What is the minimum use of specialists appropriate for children with asthma? How does that vary with history of ED or hospital use? a. When and how does the use of specialists become a marker for higher or lower quality of care? 5. What evidence is there regarding the nature of the PCP practice for children with asthma? For example, the level of continuity with individual clinicians vs. practices, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc.

IIC. **↑Asthma education:** **a. increases recognition of symptoms >** **b. ↑Management skills**	Enhancing what patients or their families know about asthma may be an important tool to improve care for children with asthma. The likely first effect of such education is to enhance the capacity of a caregiver to identify what symptoms may relate to asthma. This could conceivably increase utilization of both PCP and ED services if this were to increase the caregiver's perceived need for care for their child's asthma. With a more sophisticated understanding, including having a valid asthma action plan and understanding how to use it, ED care may be reduced and PCP care for asthma may be reduced, as symptoms are less frequent and parents are more competent to manage them when they arise.	1. What are metrics or processes regarding the quality of asthma care? Is it drug ratios (i.e. proportion of prescriptions filled that are for rescue vs control medications), asthma action plan, , capacity of PCP office, relationship to PCP practice, or other specific bundles of care, etc? 2. What constitutes "perfect care"/"best practice" for any specified type of patient? 3. What do we know about the impact of asthma education programs on quality of care, outcomes of care, or utilization of care? Define utilization of care as including: a. PCP utilization, b. ED utilization, c. Referral/specialist utilization, d. Non physician care team member utilization, e. Medication usage, f. Hospitalizations, and/or g. Other care utilization areas to consider? Examples may include functional status, quality of life elements, spirometry, role functioning. 4. What is the diversity of asthma education programs and what are the differences in quality of care/outcomes/utilization of care associated with differences? 5. Does referral to an asthma specialist impact quality of care, utilization of care and asthma outcomes? 6. Does referral to a social worker impact utilization of care and asthma outcomes? 7. (Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes? 8. What are desirable roles and effectiveness of interventions that extend beyond the healthcare system, such as reducing pollution, focusing on environmental justice, housing, dust mites, etc.? 9. How does organization and capacity of the practice setting influence the delivery of asthma management education?

Construct III:	Adequacy of PCP practice site to handle acute exacerbations of chronic disease and/or acute illnesses	
Concept	**Implications (Lay Statement)**	**Lit Review Questions**
IIIA. **↑Primary care capacity:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)** **IIIA.2** **SUBCONSTRUCT:** **↑Accessibility:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	In general, enhanced capacity may affect a patient's access to care. Capacity can refer to patient services that make it easier for a patient to receive timely care, such as location or hours of offices, to the ability to triage phone calls in a timely and effective way, or may include the materials and services present within an office (e.g. the presence of a treatment room, the capacity to deliver oxygen, nebulizers, etc.) Such capacity may be limited or enhanced by staffing, space, the ability to safely transport someone from the office to a hospital, etc. If PCP office capacity is optimized, ED visits may be reduced as acute and mundane conditions can be managed in a PCP setting. Subsequently, increased capacity of the entire PCP support network will increase number of PCP visits.	1. What do we know about access to the PCP's office as a place to manage asthma, and the subsequent capacity of a PCP and the diversity of practice settings? Additionally, how do we measure capacity and, its impact on QoC, processes of care, asthma outcomes, asthma specific processes and utilization? How do these factors impact ED use or other outcomes? a. In general: i. PCP/specialist ratio in a plan or PCP/child ratio ii. PCP time spent in visit (incl. minutes per sick, well-child, asthma management visit) iii. Nature of training activities iv. How long does it take to schedule a visit (incl. asthma (chronic), acute, follow-up visit) v. Office hours and visit flexibility (incl. after hours coverage, office consult, meet in ED) vi. Phone capabilities: (incl. answering capacity, putting on hold, returning calls, after hours phone service) vii. Level of implementation of patient centered medical home/chronic care model, eg i. Use of registries ii. Standardized tools for measurement iii. Case management iv. Group visits or other education, etc b. Specifically, ability to manage acute dx in office, which includes: i. Do they have a treatment room or capacity to use a room as a treatment room? ii. Do they offer rescue treatments (e.g. nebulizers, spacers)? iii. Can they measure oxygen saturation?

iv. Do doctors feel comfortable with acute asthmatic in office?
v. Can they take time to manage an acute pt in their office?
vi. Do they have safe and rapid transport to a hospital (how long?)

2. Availability and accessibility of offices (incl. office hours, geographic distribution)
   a. What do we know about linguistic capabilities in the PCP setting influencing use of the ED?
   b. What do we know about proximity of the PCP office to public transit on the utilization of the ED?
3. What do we know about the impact of variations in patterns of care/practice, use of modalities, and/or and receipt of well-child care on asthma management or outcomes (eg ED use)? Does Immunization status reflect on t eh capacity of the PCP, on the state of the child, or on other factors that may relate to asthma outcomes? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?
4. Are children with more WCC visits less likely to use the ED for acute visits? children who are UTD on their immunizations?
5. What literature is there on the relationship between pediatric ED use and other measures of asthma exacerbation/outcomes?
6. What do we know about variability of capacity and management of mundane conditions (e.g. OM, URIs, pharyngitis), office to ED ratios?
7. What do we know about variability of capacity and management of acute conditions requiring interventions (e.g. asthma)?
8. To what extent does ED capacity increase use of ED services? Do hospitals advertise ED services, have fast track for mundane conditions, etc?
9. To what extent does ED have capacity to provide primary care, routine immunizations, etc? How is that built into policies and

	protocols? 10. At what age does the PCP start meeting alone with child? Time spent in visit? 11. To what extent and at what age do PCP's involve children in self-management and does it vary?

IIIB. **↑Relationship with PCP:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	Improved relationship with PCP may increase visits to your PCP and decrease ED visits, for both acute and mundane conditions. A good relationship may lead to greater trust and adherence to recommendations (both WCC and asthma care) and drive a preference for seeking care by the PCP over seeking care in another environment. In general, we are referring to relationship of caregiver with PCP and their office staff. We recognize the importance of the relationship of PCP's with patients as well; when the relationship between the PCP and the child rather than caretaker is emphasized in research, we'd like to capture that as well.	1. What exists regarding measuring the quantity and quality of the relationship with PCP? Specifically: a. What's the variation and does it matter? b. How is it measured? c. What do we know about patient experience of care, especially as it relates to relationship with clinicians/PCP d. To what extent is quality of relationship expressed in terms of caregiver vs. child relationships and how does this change with age of child or longevity of connection to a PCP? 2. What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care? Define supplemental services as: a. Electronic educational/reminder tools (incl. social media) b. Telephone educational/reminder tools c. Print materials (e.g. educational brochures) d. Disease management, demand management, or other type programs e. Other services to consider? Measure quality, utilization of care should include at least : a. ED visits b. PCP visits 3. How does role of child in self care/management tie into these issues?

Construct IV: The connectedness of care in the primary care and ED setting – before, during, and after of the ED visit

Concept	Implications (Lay Statement)	Lit Review Questions
IV. (Descriptive) **Enhanced integration of ED care of asthma with routine care will have better outcomes**	If primary care is generally pretty good, then the ED visit should be an extraordinary event. In such cases the PCP alerting the ED to current management and the ED assuring appropriate follow up with the PCP is important. In cases where primary care is of lower quality or more variable, the ED visit may enhance the long term management of the child with asthma. And we need to assess this. One of the ways it might do so is to construct an asthma management plan that is then followed by the PCP. Another way is to connect a child without adequate primary care to primary care, especially to someone who is competent to manage the asthma.	1. What evidence supports that ED visits for asthma are most effective when visit is followed by a visit to the PCP? 2. Do utilization patterns in both the ED and primary care setting change following ED visits? 3. Is an effective/more effective use of medications seen following an ED visit? 4. Does the identification of a primary care provider improve outcomes of an ED visit (including patterns of care utilization)? 5. Is pre or intra visit communication with the primary care provider associated with better outcomes? How often does this occur? Are there systematic differences regarding those for whom this does and does not occur? 6. Are ED visits for asthma routinely associated with some form of communication or linkage with PCP? Does that result in better outcomes?

Construct V: Equity is a value in asthma care

Concept	Implications (Lay Statement)	Lit Review Questions
V. (Descriptive) **Equity is a critical construct of quality for children with equity**	Systematic differences in the frequency or nature of ED visits for asthma on the basis of race, ethnicity, family make-up, income/economic status, specifics of insurance status, presence or absence of comorbid special health care needs, etc represents decrements in quality that our measures should identify.	1. Does the literature indicate systematic or predictable differences in the frequency or nature of asthma care for children as it relates to ED visits for asthma that may be interpreted as representing inequitable structures, processes, outcomes, experiences with, or coordination of care? 2. What do we know about how social determinants and diagnosis and management of asthma and its outcomes, specifically as it relates to use of ED? 3. What do we know about the extent to which use of the ED for children with asthma that relates to the external physical and social environment?

Proposed Research Questions

Asthma- We propose to prioritize our Asthma Construct Table, to the following questions:

Acronyms

PCP: Primary Care Provider

ED: Emergency Department

WCC: Well-child care

Baseline Question (for Questions 1, 2 and 3 below):
When asthma care is evaluated, how is the population of asthma care recip n level? What are specific implications of how you identify patients with asthma, i o specifying the denominator of children with asthma? What are practical a ying asthma at the population level? How do the answers to these questions differ between adults and children?

Question 1 (Construct IIA.2):

For children with asthma, what do we know about asthma management? How is management of asthma described and measured? This includes who (PCP, asthma specialist, ED, etc) primarily manages it as well as who provides it. What are the patterns of care and what do we know about how use of the ED varies as a result of various approaches to management?

- **Question 1a (Construct IIB.3):**
  Specifically, have any of these patterns of visits or medication use or other characteristics of care been used as markers of well or poorly delivered primary care for asthma for children and/or adults?

Question 2 (Construct IIB.5):

How has varying asthma care for children been described on the basis of characteristics of the PCP offices or practices? For example, are they characterized by the level of continuity between individual clinicians, the level of conntiuity with any provider in the practice, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc?

- **Question 2a (Construct IIIA.3):**
  What do we know about the impact of variations in patterns of care/practice, use of treatment modalities, and/or receipt of well-child care on asthma management or outcomes (e.g. ED use)? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?

Question 3 (Construct IIC.7):

(Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes both within context of a primary care practice or in other clinical settings?

- **Question 3a. (Construct IIIB.2):**

What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care?

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.8.2. Provide the citation and summary for each piece of evidence.

Our approach to developing this measure stems from several vibrant and scientifically sound traditions. We first discuss research involving the soundness of our data sources, which include both administrative data to identify cases (and a fraction of numerator qualifications) and chart review (medical record audit) to confirm some denominator inclusions and to identify most numerator inclusion. This is a generally accepted and standard approach with acceptable reliability.

Brook and Davies [1] trace the early history of quality measurement and remind us of the importance of medical chart audit as an approach to quality measurement. Lohr and Brook at RAND and Roos in Manitoba, Canada pioneered the use of electronically-available administrative data (generated by routine health care operations, such as billings) as proxies for health care processes. Administrative data carefully used reduces burden of quality measurement. [2-6]

As the National Committee for Quality Assurance (NCQA) developed the Healthcare Employee Data Information Set (HEDIS) as the de facto measurement system for managed care, attention turned to the use of administrative data for routine performance measurement. Research demonstrated that administrative data could have a role in producing quality measures, with augmentation by chart review often necessary. Administrative data are not typically sufficient for detailed clinical assessment. [7-11] HEDIS developed a hybrid approach, using administrative data and chart review, which this measure borrows heavily from. [12, 13]

We have used rigorous and transparent methods [14] to assemble a national expert panel that included pediatricians, family physicians, pediatric and general emergency room specialists, a pediatric pulmonologist and a pediatric allergist from practices and medical schools around the country. This work was conducted in collaboration with national clinical societies (AAP, AAFP) and CAPQuaM's diverse other partner organizations, including NY State DoH/Medicaid. NCQA is an important technical consultant and partner. The specific criteria that we operationalize in this measure were all rated by the expert panel with a median score of 8 or 9 on a 9 point scale (9 high) as circumstances for which the ED is an appropriate level of care. The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children.

Select references documenting other aspects of performance gap, and supporting our process and data sources are also noted (15-35).

1. Brook, R.H. and A. Davies-Avery, *Mechanisms for Assuring Quality of US Medical Care Services: Past, Present, and Future*. 1977: Santa Monica, CA.

2. Brook, R.H. and K.N. Lohr, *Efficacy, effectiveness, variations, and quality. Boundary-crossing research.* Med Care, 1985. **23**(5): p. 710-22.

3. Lohr, K.N., *Use of insurance claims data in measuring quality of care.* Int J Technol Assess Health Care, 1990. **6**(2): p. 263-71.

4. Roos, L.L., et al., *Using Administrative Data Banks for Research and Evaluation: A Case Study.* Evaluative Research 1979. **236**(3).

5. Roos, L.L., Jr., J.P. Nicol, and S.M. Cageorge, *Using administrative data for longitudinal research: comparisons with primary data collection.* J Chronic Dis,1987. **40**(1): p. 41-9.

6. Roos, L.L., Jr., et al., *How good are the data? Reliability of one health care data bank.* Med Care, 1982. **20**(3): p. 266-76.

7. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.

8. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p.

199-207.

9. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p.

813-25.

10. Angier, H., et al., *Variation in outcomes of quality measurement by data source.* Pediatrics, 2014. **133**(6): p. e1676-82.

11. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.

12. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.

13. NCQA. *National Committee for Quality Assurance.* [cited 2014 7/30/14];

Available from: http://www.ncqa.org/.

14. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.

15. DeAngelis, C., P. Fosarelli, and A.K. Duggan, *Use of the emergency department by children enrolled in a primary care clinic.* Pediatr Emerg Care, 1985. **1**(2): p.61-5.

16. Talreja, N., et al., *Modifiable factors associated with severe asthma exacerbations in urban patients.* Ann Allergy Asthma Immunol, 2012. **109**(2): p.128-32.

17. Auger, K.A., et al., *Medical home quality and readmission risk for children hospitalized with asthma exacerbations.* Pediatrics, 2013. **131**(1): p. 64-70.

18. Ducharme, F.M., et al., *Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control.* Am J Respir Crit Care Med, 2011. **183**(2): p. 195-203.
19. Farber, H.J., *Optimizing maintenance therapy in pediatric asthma.* Curr Opin Pulm Med, 2010. **16**(1): p. 25-30.
20. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, *Relationship between pediatric primary provider visits and acute asthma ED visits.* Pediatr Pulmonol, 2007. **42**(11): p. 1041-7.
21. Parchman, M.L. and S. Culler, *Primary care physicians and avoidable hospitalizations.* J Fam Pract, 1994. **39**(2): p. 123-8.

22. Prevention, C.f.D.C.a. *Home-based Multi-trigger, Multi-component interventions*. 2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
23. Price, J.H., et al., *Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States.* Biomed Res Int, 2013. **2013**: p. 1-12; Available in open access http://dx.doi.org/10.1155/2013/787616
24. Homer, C.J., et al., *Does quality of care affect rates of hospitalization for childhood asthma?* Pediatrics, 1996. **98**(1): p. 18-23.
25. Finkelstein, J.A., et al., *Quality of care for preschool children with asthma: the role of social factors and practice setting.* Pediatrics, 1995. **95**(3): p. 389-94.
26. Flores, G., et al., *Keeping children out of hospitals: parents' and physicians' perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided.* Pediatrics, 2003. **112**(5): p. 1021-30.
27. Knudson, A., et al., *Disparities in pediatric asthma hospitalizations.* J Public Health Manag Pract, 2009. **15**(3): p. 232-7.
28. Bindman, A.B., et al., *Preventable hospitalizations and access to health care.* Jama, 1995. **274**(4): p. 305-11.
29. National Heart, Lung and Blood Institute. *Asthma Guidelines*. 2011 February 2011 [cited 2014 7/30/2014].
30. *A Matter of Urgency: Reducing Emergency Department Overuse* in *A NEHI Research Brief* 2010, New England Healthcare Institute.
31. Martin, B.C., *Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital.* J Health Care Finance, 2000. **27**(2): p. 51-65.
32. Finkelstein, J.A., et al., *Comparing asthma care for Medicaid and non-Medicaid children in a health maintenance organization.* Arch Pediatr Adolesc Med, 2000. **154**(6): p. 563-8.
33. Owens, P.L., et al., *Care of children and adolescents in U.S. hospitals*, in *HCUP Fact Book No. 4*. Agency for Healthcare Research and Quality: Rockville: MD.
34. Pearson, W.S., et al., *State-based Medicaid costs for pediatric asthma emergency department visits.* Prev Chronic Dis, 2014. **11**: p. E108.
35. Taubman, S.L., et al., *Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment.* Science, 2014. **343**(6168): p. 263-268.

The appropriate use criteria were derived from a set developed by an expert panel who synthesized the literature and their expert opinion into explicit criteria using the RAND/UCLA appropriateness method. Criteria that were rated 8 or 9 by the panel were included for this measure. The criteria set includes:

1) Hospitalization directly from the ED;

2) Documented physical findings consistent with respiratory distress, including:

   a) Labored breathing with retractions and/or evidence of accessory muscle use;

   b) Markedly decreased breath sounds;

3) O_2 saturation level less than 90 percent on percutaneous assessment;

4) An ABG obtained (or ordered);

5) Consultation ordered and obtained with a pulmonologist asthma specialist, an order of an arterial blood gas (ABG), or a consult with a pulmonary or asthma specialist.

6) Parent/caregiver referred to the ED after evaluation from the PCP or other office/clinic;

7) Parent/caregiver report of administering two or more doses of inhaled rescue medications without meaningful clinical improvement;

8) Parent/caregiver report that the child was in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan; or,

9) Parent/caregiver report of a rapid and life-threatening deterioration after a similar prior episode. This criterion is not included in the specifications for this measure.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_Evidence_submission_form_-_Appropriate_emergency_Final.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Asthma is one of the most common indications for emergency department (ED) visits by children. (1-3) AHRQ's Healthcare Cost and Utilization Project (HCUP) data from the Nationwide Emergency Department Sample (NEDS) found that in 2012, children between 1 and 17 years old had more than 1,895,000 ED visits for asthma with almost 10% resulting in hospitalization.

Evidence suggests that ED visits and hospitalizations in children with asthma vary systematically by how well-equipped that community is to provide primary care, and by the quality of primary care delivered. (4, 5) There is widespread literature illustrating that ED visits and hospitalizations are each undesirable utilization outcomes from poorly managed asthma. There is not a large literature that assesses whether or not pediatric ED visits were appropriate. (6 -10)

A body of literature has explored the value and feasibility of measuring the appropriateness of medical activities using data available in the medical record. (11-14) Early work in adults included assessment of hysterectomy, carotid endarterectomy and cardiac interventions. An independent research project brought the construct of appropriateness to children (15), while Kleinman and colleagues were the first to assess the appropriateness of specific pediatric procedures. (16, 17) A later study demonstrated the feasibility of medical record data for such an assessment. (18) DeAngelis pioneered studies of what constitutes a good reason to use the ED. (6) All of these studies used a definition of appropriateness that compared benefit to likely risk without specific consideration of costs. The need for more studies looking for overuse was recently reviewed. (19) RAND type Delphi panels are accepted around the world as a method for developing criteria to assess appropriateness. (20-22)

Research demonstrates that:

•ED visits are an important issue for child health insurers, including Medicaid, with clinical and financial consequences;
•An overcrowded primary care system contributes to ED use for non-emergent and even non-urgent conditions.
•Pediatric hospitalizations for asthma vary by primary care availability and quality
•ED visits are common for children with asthma, including those in Medicaid
•Assessment of appropriateness using information in the medical record is a well-established and validated method that has been successfully applied to children.

The literature suggests that a measure that assesses whether or not the ED is an appropriate level of care for a child with asthma at the time that they present has intrinsic value. Such a measure would:

•Characterize the process of care in a way that assesses whether a particular ED visit represents overuse
•Allow the outcomes of asthma care to be better characterized in a manner that describes performance and promotes targeted

improvement. Inappropriate ED visits represent failures of primary care delivery, availability and/or access. Appropriate visits may represent a failure to control asthma. These have distinct and distinguishable meanings that contribute to the understanding of the quality of asthma care.
•Measuring the quality of asthma care requires assessment of multiple factors. This appropriateness measure helps plans, purchasers, and society to understand the implication of asthma ED visits as outcomes of asthma care. The implications herein is that understanding what is better or worse care requires looking at various factors and not simply a higher or lower appropriateness score. The understanding of this measure is enhanced by considering whether the rate of undesirable outcomes (ED visits and hospitalizations) is high or low and whether other measures of primary care availability and access or asthma quality suggest high levels of performance or not..

An abstract describing the proposed measure was peer-reviewed and subsequently presented to a national audience at AcademyHealth 2014 Annual Research Meeting in San Diego in the "Measuring the Safety, Quality, and Value" section. Feedback was positive regarding the methods, measures, ethics, and importance of this measure.

Research evidence supports the importance and need for our proposed measure that assesses whether the ED represents an appropriate level of care for children with asthma who are seen in the ED.

1.Kharbanda, A.B., et al., Variation in resource utilization across a national sample of pediatric emergency departments. J Pediatr, 2013. 163(1): p. 230-6.
2.Adams, J.G., Emergency department overuse: Perceptions and solutions. JAMA, 2013. 309(11): p. 1173-1174.
3.Institute, N.E.H., A Matter of Urgency: Reducing Emergency Department Overuse. Research Brief, 2010(March).
4.Perrin, J.M., et al., Variations in rates of hospitalization of children in three urban communities. N Engl J Med, 1989. 320(18): p. 1183-7.
5.Perrin, J.M., et al., Primary care involvement among hospitalized children. Arch Pediatr Adolesc Med, 1996. 150(5): p. 479-86.
6.DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p.61-5.
7.Berns, S.D., et al., Appropriate use of a pediatric emergency department: is the pediatrician called before the visit? Pediatr Emerg Care, 1994. 10(1): p. 13-7.
8.Rudowitz, R., A Look At CBO Projections For Medicaid and CHIP, in The Kaiser Commission on Medicaid and the Uninsured. 2014, The Henry J. Kaiser Family Foundation
9.Taubman, S.L., et al., Medicaid Increases Emergency-Department Use: Evidence from Oregon´s Health Insurance Experiment. Science, 2014. 343(6168): p. 263-268.
10.Smulowitz, P.B., et al., Increased Use of the Emergency Department After Health Care Reform in Massachusetts. Ann Emerg Med, 2014.
11.Brook, R.H., et al., A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care, 1986. 2(1): p. 53-63.
12.Park, R.E., et al., Physician ratings of appropriate indications for six medical and surgical procedures. Am J Public Health, 1986. 76(7): p. 766-72.
13.Fitch, K., et al., The RAND/UCLA Appropriateness Method User´s Manual. 2001 RAND.
14.Kosecoff, J., et al., The appropriateness of using a medical procedure. Is information in the medical record valid? Med Care, 1987. 25(3): p. 196-201.
15.Kemper, K.J., Medically inappropriate hospital use in a pediatric population. N Engl J Med, 1988. 318(16): p. 1033-7.
16.Kleinman, L.C., et al., The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. Jama, 1994. 271(16): p. 1250-5.
17.Kleinman, L.C., E.A. Boyd, and J.C. Heritage, Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians. Jama, 1997. 278(6): p. 497-501.
18.Keyhani, S., et al., Electronic health record components and the quality of care. Med Care, 2008. 46(12): p. 1267-72.
19.Keyhani, S. and A.L. Siu, The underuse of overuse research. Health Serv Res, 2008. 43(6): p. 1923-30.
20.Bernstein, S.J., et al., The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium. Jama, 1993. 269(18): p. 2398-402.
21.Taylor, A.J., et al., ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 Appropriate Use Criteria for Cardiac Computed Tomography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. J Cardiovasc Comput

Tomogr, 2010. 4(6): p. 407.e1-33.
22.Basger, B.J., T.F. Chen, and R.J. Moles, Validation of prescribing appropriateness criteria for older Australians using the RAND/UCLA appropriateness method. BMJ Open, 2012. 2(5).

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

In testing we found that for children age 2-5, 181 of 335 (54.3%) were deemed appropriate, with the breakdown of reasons for appropriateness presented in the Testing Form 2b.2.2. Other age groups found that children 6-11, 209 of 447 (43.8%) ED visits were appropriate, while for adolescents aged 12-18, 165 of 341 (48.4%) visits were appropriate. These numbers were sufficient to identify statistically significant differences in the proportion that were appropriate between age groups, among racial/ethnic groups, and within age group among racial/ethnic groups. These data demonstrate that the specified sample size is sufficient to find meaningful differences between groups at the various specified levels. In our work, validating and testing the measure for the rate of appropriateness, we have demonstrated the capacity to identify the included events (ED visits and hospitalizations) using administrative data and our specifications for identifiable asthma. That aspect of testing was conducted using state wide data from the NY State Medicaid Managed Care Program.

As noted in the evidence form, we also incorporate by reference work done by our partner NCQA that demonstrate the capacity of administrative data to identify a population with asthma

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

In her seminal article nearly three decades ago, DeAngelis included an asthma attack as an appropriate indication for use of the ED.[1] As a common chronic illness characterized by remissions and potentially preventable exacerbations undesirable utilization outcomes for asthma have been a frequent target for measurement for three decades. Reducing the relative number of ED visits during the care for asthmatic children remains a high priority on the national agenda. The universal delivery of optimal asthma care has the potential to lower costs and improve quality of life. Understanding which ED visits represent failures of clinical prevention and which instead represent a mismatch of service level to clinical need can help to move these goals forward. The submitted measure is a step in this direction.

ED visits for asthma can be reduced through both enhanced access to care and through better quality of care. The NIH's National Asthma Education and Prevention Program Guideline has been shown to reduce the frequency of breakthrough asthma and of ED visits and hospitalizations when implemented. The literature points to two general characteristics of asthma care delivery systems that correlate with ED utilization. One is the effective use of preventive and routine care measures, such as multidisciplinary practice or a medical home model, the presence of an asthma action plan, the use of controller medications supplemented by judicious use of rescue medications. [2-6] The other is the availability of primary care or urgent care visits as a step before ED use in the context of either a general pediatric or an asthma specialty practice. [6, 7] Conversely, a lack of comprehensive asthma care, which includes primary and secondary prevention schemas, and a lack of available urgent care services are both commonly cited as reasons for preventable ED visits. It has been demonstrated that the children who used the ED underutilized primary care services [6] and it has also been demonstrated that interventions that attempt to provide comprehensive, multidisciplinary care are able to decrease ED utilization for asthma care.[8] We acknowledge that environmental management and control is a nonclinical opportunity to improve the quality of life for children with asthma and to reduce health care utilization, but do not focus on these issues in this submission.

High rates of asthma visits to the ED suggest widespread deficiencies in asthma care. The literature shows that lack of proper asthma care is disparate with minority children bearing undue burden. [9-11]

The literature also presents different perspectives on appropriate use of the ED for pediatric asthma. Pediatric asthma is one of the leading conditions when it comes to potentially avoidable ED visits. [12] Asthma has been classified both as an avoidable hospitalization condition (AHC) and as an ambulatory care sensitive condition. This describes that a meaning proportion of ED visits or hospital admissions could have been avoided with proper outpatient care. [12, 13] Poor outpatient care can be an outcome of a number of variables. As noted, the availability of primary care can reduce such inappropriate and costly visits. [7, 12, 14 -17]

Assessing the extent to which ED use for asthma is appropriate can inform health policy, manpower planning, and clinical quality

improvement activities. It can help to answer the question of how much of ED use potentially may be prevented by better management of the underlying asthma, versus how much requires other, process or structural improvements to reduce use of the ED when a lower level of care would meet the clinical needs of the child. Refractory asthma or those with unavoidable environmental exposures leading to an acute exacerbation requiring medical care are likely to be identified as appropriate, reminding us that NOT all asthma ED visits are preventable even with optimal care.

With a better understanding of ED use, health care organizations and policy makers could develop better informed approaches to optimizing services for children with asthma. And hopefully children and their families may increasingly be spared the inconvenience, risk, and costs of ED visits for asthma.

References:

1. DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p.61-5.
2. Talreja, N., et al., Modifiable factors associated with severe asthma exacerbations in urban patients. Ann Allergy Asthma Immunol, 2012. 109(2): p.128-32.
3. Auger, K.A., et al., Medical home quality and readmission risk for children hospitalized with asthma exacerbations. Pediatrics, 2013. 131(1): p. 64-70.
4. Ducharme, F.M., et al., Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control. Am J Respir Crit Care Med, 2011. 183(2): p. 195-203.
5. Farber, H.J., Optimizing maintenance therapy in pediatric asthma. Curr Opin Pulm Med, 2010. 16(1): p. 25-30.
6. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, Relationship between pediatric primary provider visits and acute asthma ED visits. Pediatr Pulmonol, 2007. 42(11): p. 1041-7.
7. Parchman, M.L. and S. Culler, Primary care physicians and avoidable hospitalizations. J Fam Pract, 1994. 39(2): p. 123-8.
8. Prevention, C.f.D.C.a. Home-based Multi-trigger, Multi-component interventions.
2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
9. Price, J.H., et al., Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States. Biomed Res Int, 2013. 2013: p. 1-12; Available in open access http://dx.doi.org/10.1155/2013/787616
10. Homer, C.J., et al., Does quality of care affect rates of hospitalization for childhood asthma? Pediatrics, 1996. 98(1): p. 18-23.
11. Finkelstein, J.A., et al., Quality of care for preschool children with asthma: the role of social factors and practice setting. Pediatrics, 1995. 95(3): p. 389-94.
12. Flores, G., et al., Keeping children out of hospitals: parents´ and physicians´ perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided. Pediatrics, 2003. 112(5): p. 1021-30.
13. Knudson, A., et al., Disparities in pediatric asthma hospitalizations. J Public Health Manag Pract, 2009. 15(3): p. 232-7.
14. Bindman, A.B., et al., Preventable hospitalizations and access to health care. Jama, 1995. 274(4): p. 305-11.
15. National Heart, Lung and Blood Institute. Asthma Guidelines. 2011 February 2011 [cited 2014 7/30/2014].
16. A Matter of Urgency: Reducing Emergency Department Overuse in A NEHI Research Brief 2010, New England Healthcare Institute.
17. Martin, B.C., Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital. J Health Care Finance, 2000. 27(2): p. 51-65.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Race/Ethnicity

Our medical chart audit found that the measure varies by race/ethnicity. Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (46.8%), p=.002. Black children showed a trend toward more questionable use compared to all other children (53.6% questionable vs 48.7%, p=.10). Overall, Blacks had an appropriate use rate of 51.3%, Whites 56.5%, Hispanics 44.1% and other races 45.2%. For ages 2-5, Blacks had an appropriate use rate of 57.1%, Whites 63.6%, Hispanics 50.9%, and other races 51.9%. For ages 6-11, Blacks had an appropriate use rate of 49.3%, Whites 50.0%, Hispanics 46.3% and other races 39.8%. For ages 12-18, Blacks had an appropriate use rate of 49.3%, Whites 66.7%, Hispanics 46.3% and other races 46.7%. Chi-square analysis confirms that these differences are statistically significant.

Insurance Status

Overall, the appropriate use rate for Medicaid patients was 46.3%, Private insurance 59.0%, Uninsured patients 38.6% and other forms of insurance (military and Worker's comp) 55.0% (p=.005). Within the age strata, for ages 2-5 the appropriate use rate for Medicaid patients was 53.9%, Private 67.4%, Uninsured 40.9% and other was 20%. For ages 6-11, the appropriate use rate for Medicaid patients was 41.5%, Private 57.7%, Uninsured 35.7% and other was 52.6%. For ages 12-18, the appropriate use rate for Medicaid patients was 46.1%, Private 54.5%, Uninsured 42.1% and other was 68.8%. Chi-square analysis demonstrates the presence of statistically significant differences.

Socioeconomic Status
The measure is specified to be stratified in 2 ways to assess aspects related to socioeconomic status: Public versus Commercial Insurance, and by 5 strata defined by the percent of the population in poverty in their county of residence.

Rurality/Urbanicity
These measures are specified to be reported by Urban Influence Codes (UIC), which have been developed by the USDA based on a number of criteria to describe the levels of urbanicity and rurality. This is intended not only to report within plan differences but to allow for aggregation as appropriate. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. We recommend consideration of the aggregation schema of Bennett and colleagues at the South Carolina Rural Research Center. (2) Their aggregation scheme brings together Codes 1 & 2 as Urban; 3, 5, & 8 as micropolitan rural; 4, 6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We observe that UIC 5 might as well be aggregated with 4, 6, & 7 as an adjacent rural area. Further, while this approach to rurality does not map exactly to the population density based definition of frontier (< 6 persons per square mile) as articulated in the Affordable Care Act, use of such categories is consistent with the ACA's intent that the Secretary ask that data that are collected for racial and ethnic disparities also look at underserved frontier counties. Frontier health care may be approximated by analysis of the remote rural categories. (3)

This judgment was confirmed after CAPQuaM consulted with Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Sciences, who is heading a HRSA-funded project to develop new methods to analyze frontier health. We clarified that his work suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in sensitivity.

Those interested in care specific to large cities may wish to aggregate the rural area and analyze UIC 1 and 2 separately. Frontier health care may be approximated by analysis of the remote rural categories. (3) The New York State Medicaid data were sensitive to urbanicity with higher rates of ED utilization in the most urban areas and lowest in the most rural areas and other areas intermediate between the two.

For aggregation and as an imperfect approximation one can also group as urban (1 and 2), suburban (3-6) and rural (7-9). This is what we have used for our NY Medicaid analysis to demonstrate that variations are observed for this measure using UIC codes.

1.Kawachi I, B.L., Neighborhoods and Health. 2003, New York, NY: Oxford University Press.
2.Bennett, K.J., Olatosi B. & Probst, J.C., Health Disparities: A rural-urban chartbook. 2008, Columbia, South Carolina: South Carolina Rural Health Research Center.
3.Hart, G., Frontier/Remote, Island, and Rural Literature Review. 2012.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Performance data provided in 1b.4

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Asthma matters for pediatrics. [1-12] It is one of the most common chronic diseases in children, affecting an estimated 7.1 million children in the United States. [13] In 2011, 4.1 million children suffered from an asthma attack or episode. It is the second most common reason (after allergy) for children to be classified as having a special health care need, accounting for nearly 38.8% of such children. Pediatric asthma is more prevalent in minority populations. Lifetime prevalence rates of asthma in Hispanic and Black children are 12.4% and 15.8% respectively. [14]

AHRQ's Healthcare Cost and Utilization Project (HCUP) data from the Nationwide Emergency Department Sample (NEDS) found that in 2012, children between 1 and 17 years old had more than 1,895,000 ED visits for asthma with almost 10% resulting in hospitalization. Asthma ED visits are common across a diversity of hospital types, and in all regions of the country with more than three quarters occurring in metropolitan areas. [9-12]

With our proposed measure, CAPQuaM advances DeAngelis' seminal work (15) by implementing systematically-developed explicit criteria to assess whether or not there is information to document that the ED was the appropriate level of care for the specific presentation of a given child.

High rates of asthma visits to the ED suggest widespread deficiencies in asthma care (16-22). The literature shows that lack of proper asthma care is disparate with minority children bearing undue burden. [23-25]

The literature also presents diverse perspectives on appropriate use of the ED for pediatric asthma (21, 26-28). Availability of primary care can reduce inappropriate visits to the ED. [21, 26, 28 -42] The literature does not always make clear that a potentially preventable visit is not the same as an inappropriate or unnecessary visit – sick asthmatic children may require ED care. The capacity to describe visits as appropriate will be informative to policy makers as they consider workforce issues regarding the capacity needed to maintain adequate emergency room services.

Assessing the extent to which ED use for asthma is appropriate can inform health policy, manpower planning, and clinical quality improvement activities. It can help to answer the question of how much of ED use potentially may be prevented by better management of the underlying asthma, versus how much requires other, process or structural improvements to reduce use of the ED when a lower level of care would meet the clinical needs of the child. Refractory asthma or those with unavoidable environmental exposures leading to an acute exacerbation requiring medical care are likely to be identified as appropriate, reminding us that NOT all asthma ED visits are preventable even with optimal care.

We have previously submitted to the PQMP a measure that uses an algorithm validated by an expert panel to identify children who have asthma that had required health care services in the recent past and their asthma is sufficient that it should have been identified and managed by the health care system. Only children who have such identified asthma are considered eligible for this current measure. Previously we used 2010 and 2011 data and found that more than 196,000 such children in New York State have identifiable asthma; more than forty thousand of those children generated nearly 60,000 asthma-related ED visits in 2011. We have further submitted measures that assess proxies for linkages between the primary care and ED systems. This measure fills a gap by further distinguishing those ED visits for which one can identify in the medical record an indication that makes the ED visit an appropriate level of care and those for which such an indication cannot be identified. We call the former circumstance "appropriate" and the latter "questionable" to reflect our uncertainty about legitimate reasons for using the ED that may not be recorded routinely in the medical record (including several patient-centered reasons identified by our expert panel).

A recent RAND systematic review of non-urgent ED use lamented the lack of a standardized definition for what constitutes a non-urgent ED visit. [43] In the context of our assignment to develop measures related to "asthma ED, overuse" we have translated the RAND observation into a well-specified approach to assess whether or not the ED is an appropriate level of service for a specified child given the totality of their current circumstances. We assess this using explicit criteria developed by an expert panel incorporated into a modified RAND-UCLA Appropriateness Method.

1c.4. Citations for data demonstrating high priority provided in 1a.3

1.Organization, W.H. World Health Organization Fact Sheet #307. updated May 2011 [cited 2013 May 20]; Available from: http://www.who.int/mediacentre/factsheets/fs307/en/

2.Adams, R.J., B.J. Smith, and R.E. Ruffin, Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. Thorax, 2000. 55(7): p. 566-73.
3.Bahadori, K., et al., Economic burden of asthma: a systematic review. BMC Pulm Med, 2009. 9: p. 24.
4.Weiss, K.B., P.J. Gergen, and T.A. Hodgson, An economic evaluation of asthma in the United States. N Engl J Med, 1992. 326(13): p. 862-6.
5.Coventry, J.A., M.S. Weston, and P.M. Collins, Emergency room encounters of pediatric patients with asthma: cost comparisons with other treatment settings. J Ambul Care Manage, 1996. 19(2): p. 9-21.
6.Association, A.L. American Lung Association Fact Sheet on Asthma and Children. October 2012 [cited 2013 May 20]; Available from: http://www.lung.org/lung-disease/asthma/resources/facts-and-figures/asthma- children-fact-sheet.html#5.
7.Fuhrman, C., et al., Hospitalizations for asthma in children are linked to undertreatment and insufficient asthma education. J Asthma, 2011. 48(6): p.565-71.
8.Sawicki, G.S., et al., Uncontrolled asthma in a commercially insured population from 2002 to 2007: trends, predictors, and costs. J Asthma, 2010. 47(5): p. 574-80.
9.Manice, M., Exploring the relationship between parental shared decision-making practices and acute asthma exacerbations among children age 0-17. 2013, Icahn School of Medicine at Mount Sinai: New York, NY.
10.Cerdan, N.S., et al., Asthma severity in children and the quality of life of their parents. Appl Nurs Res, 2012. 25(3): p. 131-7.
11.Fiese, B., et al., Family climate of routine asthma care: associating perceived burden and mother-child interaction patterns to child well-being. Fam Process, 2008. 47(1): p. 63-79.
12.Okelo, S.O., et al., Emotional quality-of-life and outcomes in adolescents with asthma. J Pediatr, 2004. 145(4): p. 523-9.
13.Prevention, C.f.D.C.a., National Center for Health Statistics, National Health Interview Survey Raw Data, Analysis by the American Lung Association Research and Health Education Division using SPSS and SUDAAN software. 2011.
14.Lara, M., et al., Heterogeneity of childhood asthma among Hispanic children: Puerto Rican children bear a disproportionate burden. Pediatrics, 2006. 117(1): p. 43-53.
15.DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p. 61-5.
16.Talreja, N., et al., Modifiable factors associated with severe asthma exacerbations in urban patients. Ann Allergy Asthma Immunol, 2012. 109(2): p. 128-32.
17.Auger, K.A., et al., Medical home quality and readmission risk for children hospitalized with asthma exacerbations. Pediatrics, 2013. 131(1): p. 64-70.
18.Ducharme, F.M., et al., Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control. Am J Respir Crit Care Med, 2011. 183(2): p. 195-203.
19.Farber, H.J., Optimizing maintenance therapy in pediatric asthma. Curr Opin Pulm Med, 2010. 16(1): p. 25-30.
20.Smith, S.R., D.B. Wakefield, and M.M. Cloutier, Relationship between pediatric primary provider visits and acute asthma ED visits. Pediatr Pulmonol, 2007. 42(11): p. 1041-7.
21.Parchman, M.L. and S. Culler, Primary care physicians and avoidable hospitalizations. J Fam Pract, 1994. 39(2): p. 123-8.
22.Prevention, C.f.D.C.a. Home-based Multi-trigger, Multi-component interventions. 2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
23.Price, J.H., et al., Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States. Biomed Res Int, 2013. 2013: Article ID 787616.
24.Homer, C.J., et al., Does quality of care affect rates of hospitalization for childhood asthma? Pediatrics, 1996. 98(1): p. 18-23.
25.Finkelstein, J.A., et al., Quality of care for preschool children with asthma: the role of social factors and practice setting. Pediatrics, 1995. 95(3): p. 389-94.
26.Flores, G., et al., Keeping children out of hospitals: parents´ and physicians´ perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided. Pediatrics, 2003. 112(5): p. 1021-30.
27.Knudson, A., et al., Disparities in pediatric asthma hospitalizations. J Public Health Manag Pract, 2009. 15(3): p. 232-7.
28.Bindman, A.B., et al., Preventable hospitalizations and access to health care. Jama, 1995. 274(4): p. 305-11.
29.National Heart, L., and Blood Institute. Asthma Guidelines. 2011 February 2011 [cited 2014 7/30/2014].
30.A Matter of Urgency: Reducing Emergency Department Overuse in A NEHI Research Brief 2010, New England Healthcare Institute.
31.Martin, B.C., Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital. J Health Care Finance, 2000. 27(2): p. 51-65.
32.Finkelstein, J.A., et al., Comparing asthma care for Medicaid and non-Medicaid children in a health maintenance organization. Arch Pediatr Adolesc Med, 2000. 154(6): p. 563-8.
33.Owens, P.L., et al., Care of children and adolescents in U.S. hospitals, in HCUP Fact Book No. 4. Agency for Healthcare Research and Quality: Rockville: MD.
34.Pearson, W.S., et al., State-based Medicaid costs for pediatric asthma emergency department visits. Prev Chronic Dis, 2014. 11: p.

E108.
35.Taubman, S.L., et al., Medicaid Increases Emergency-Department Use: Evidence from Oregon´s Health Insurance Experiment. Science, 2014. 343(6168): p. 263-268.
36.Dick, S., et al., Associations between environmental exposures and asthma control and exacerbations in young children: a systematic review. BMJ Open, 2014. 4(2): p. e003827.
37.Kearney, G.D., et al., Eastern Carolina Asthma Prevention Program (ECAPP): An Environmental Intervention Study Among Rural and Underserved Children with Asthma in Eastern North Carolina. Environ Health Insights, 2014. 8: p. 27 - 37.
38.Roy, A., M.J. Downes, and J.P. Wisnivesky, Comprehensive environmental management of asthma and pediatric preventive care. Pediatr Allergy Immunol, 2011. 22(3): p. 277-82.
39.Roy, A., et al., The effects of outdoor air pollutants on the costs of pediatric asthma hospitalizations in the United States, 1999 to 2007. Med Care, 2011. 49(9): p. 810-7.
40.Roy, A. and J.P. Wisnivesky, Comprehensive use of environmental control practices among adults with asthma. Allergy Asthma Proc, 2010. 31(5): p. 72-7.
41.Roy, A. and J.P. Wisnivesky, Racial and ethnic differences in the use of environmental control practices among children with asthma. J Asthma, 2010. 47(5): p. 507-12.
42.Downes, M.J., et al., Factors associated with furry pet ownership among patients with asthma. J Asthma, 2010. 47(7): p. 742-9.
43.Uscher-Pines, L., et al., Emergency department visits for nonurgent conditions: systematic literature review. Am J Manag Care, 2013. 19(1): p. 47-59.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Prevention, Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Access, Care Coordination, Overuse

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
We currently do not have a web page. We will ensure that this measure will be publicly available.

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** FINAL_CAPQuaM_ASTHMA_ICD9_and_ICD10-635802445620975487.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date

and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The numerator is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. Distinct numerators are reported for children ages 2-5, 6-11, 12-18, and optionally, 19 - 21.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Two years of administrative data are needed for this analysis: the reporting year and the 12 months preceding the reporting year.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Children and adolescents who have a qualifying ED visit associated with asthma as the first or second diagnosis;

AND have at least one of the following:

•Disposition of the ED visit was admission to the hospital
•Documented physical findings consistent with respiratory distress, including any of the following:

o Labored breathing (including moderate or severe increased work of breathing);
o Retractions, grunting, and/or evidence of accessory muscle use;
o Markedly decreased breath sounds;

•Low oxygen (O2) saturation level (dichotomized, < 90% qualifies);
•An arterial blood gas (ABG) was obtained in the emergency department;
•The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED;
• There is clear documentation that prior to arrival in the ED any of the following occurred:

o The child was referred to the ED after evaluation by the PCP or other clinician
- note: assessment of breathing over the telephone is allowed by this criterion;
o The child received two or more doses of inhaled rescue medications without sufficient clinical improvement. Note: parental report of this criterion is acceptable. Report may have been made at triage, to the nursing staff, or by the clinician during the chief complaint or history of present illness;
o The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan. Documentation is needed that the patient/family OR physician report or the chart documents ALL of the following
- a written asthma action plan exists AND defines a "red zone" for which urgent assessment by a clinician is indicated;
- An objective assessment was made and its result was in the pre-defined red zone

These details incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator represents a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (Appendix Table 1).

Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. An overall rate

across strata is not reported.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Denominator Elements:

The presence of identifiable asthma (see table 1) is established each month from administrative data using the specified algorithm.

Descriptive definitions for being managed for identifiable asthma are as follows. Specifications follow the descriptive definitions. Identifiable asthma is present in any child who has:

- Any prior hospitalization with asthma as primary or secondary diagnosis; or,
- Other qualifying events, all ages:

o Three or more ambulatory visits with diagnosis of asthma or bronchitis,
OR
o Two or more ambulatory visits with a diagnosis of asthma and/or bronchitis AND one or more asthma-related prescriptions

- OR For children older than five who have an ED visit for asthma (as first or second diagnosis) in the reporting month and prior to the reporting month who have had:

o One or more prior ambulatory visits with asthma as the primary diagnosis after the fifth birthday, OR

o Two or more ambulatory visits after the fifth birthday with asthma as a diagnosis, OR
o One ambulatory visit with asthma as a diagnosis AND at least one asthma-related prescription, both occurring after the fifth birthday OR
o Two or more ambulatory visits with a diagnosis of bronchitis after the fifth birthday

For eligibility purposes, asthma-related medicine means long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti- asthmatic combinations, methylxanthines (alone or in combination), and/or mast cell stabilizers. See below further regarding this specification. Note that leukotriene modifiers and short term beta agonists are excluded for the purpose of establishing identifiable asthma. Data from the year prior to the reporting year are used, as well as all months prior to the reporting month in the reporting year (see Appendix Figure 1).

All events in the administrative data should be associated with a date of service.

Eligibility should be obtained using the month by month algorithm described herein and illustrated in Figure1, which is a fundamental component of this description. The analysis should be conducted on a month by month basis as described herein:

Within the group of children who meet the criteria for identifiable asthma, identify and maintain a unique patient identifier, age, and all stratification variables.

Determine eligibility for each patient, as of the last day of the month prior to the reporting month.

For example, if the goal is to report for January 2011, first identify children with identifiable asthma (above), and analyze all of calendar year 2010 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2010.

Next, for February analyze all of calendar year 2010 AND January 2011. Continuous enrollment criterion requires that the child was enrolled in December
2010 and January 2011.

Repeat this progression monthly so that for December, one would identify children with identifiable asthma and analyze all of calendar year 2010 AND January through November 2011 when doing so. Continuous enrollment criterion requires that for December the child was enrolled in October 2011 and November 2011.

See Figure 1 in Appendix.

Develop Denominator sample according to Appendix Figure 2 and consistent with the instructions in sections S.18 and S.20.

Codes used for definitions are specified in Appendix Table 1 and summarized herein:

Hospitalization:

CPT Codes: (Any)

CPT 99238 CPT 99232
CPT 99239 CPT 99233
CPT 99221 CPT 99234
CPT 99222 CPT 99235
CPT 99223 CPT 99236
CPT 99356 CPT 99218
CPT 99357 CPT 99219
CPT 99231 CPT 99220

Or Revenue Codes: (Any)

0110	0133
0111	0134
0112	0137
0113	0139
0114	0150
0117	0151
0119	0152
0120	0153
0121	0154
0122	0157
0123	0159
0124	0200
0127	0201
0129	0202
0130	0203
0131	0204
0132	0206

Emergency Department Visits

CPT Codes: (Any)

CPT 99281	CPT 99284
CPT 99282	CPT 99285
CPT 99283	

Or Revenue Codes: (Any)

0450 Emergency Room

0451 Emergency Room: EM/EMTALA
0452 Emergency Room: ER/Beyond EMTALA
0456 Emergency Room: Urgent Care
0459 Emergency Room: Other Emergency Room
0981 Professional Fees (096x) Emergency Room
981 Professional Fees emergency room

Office Visits(Any)

CPT 99201 CPT 99211
CPT 99202 CPT 99212
CPT 99203 CPT 99213
CPT 99204 CPT 99214
CPT 99205 CPT 99215

Diagnosis of Asthma

ICD-9 Codes:
All codes beginning with 493
Please see the Excel spreadsheet on s.2.b. for detailed list of ICD9 codes and specified list of ICD 10 codes.

Filled Prescriptions for Asthma-related Medications
Use NCQA NDC list (ASM-C_DASM-C_final_2012, found by clicking through at (http://www.ncqa.org/HEDISQualityMeasurement/HEDISMeasures/HEDIS2012/HEDIS2012FinalNDCLists.aspx) Eliminate medications in the following2 categories: leukotriene modifiers, short-acting inhaled beta-2 agonists). May use equivalent updated lists when provided by NCQA.

Please note Figures 1 and 2 and Table 1 in the attached Appendix are considered INTEGRAL to these specifications and are not optional.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema. Identifiable asthma is defined is section S.9.

At the discretion of the accountability entity, the denominator may be restricted to children 2-18.

These details incorporate ICD-9 codes only. For the specified ICD-10 codes and a detailed listing of ICD 9 codes see attached spreadsheet in S2.b.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Denominator Exclusions
1) Children with concurrent or pre-existing:
a. Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 code: 496);
b. Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03,277.09) ;
c. Emphysema diagnosis (ICD-9 code 492xx)
2) Children without identifiable asthma as defined in S.9 by the month before the ED visit
3) Outside of specified age range
4) Events occurring in patients who have not been enrolled in the reporting plan for at least two consecutive months before the index reporting month (a total of 3 consecutive months, including the reporting month).

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

Specifications for this measure requires stratification by age group. Several additional stratifications are optional but may be required by the accountability entity. These variables include race/ethnicity, rurality/urbanicity and county level of poverty.

Stratify by age group (use age at month of qualifying event):
•Age 2-5 years (second birthday to the day before the 6th birthday);
•Age 6-11 years (sixth birthday to the day before the 12th birthday);
•Age 12-18 years (twelfth birthday to the day before the 18th birthday); and
•Age 19-21 years (nineteenth birthday to the day before the 21st birthday).

Age strata are to be reported distinctly and not combined.

Optional stratifications require data elements such as:
•Race/Ethnicity
•Insurance type (Public, Commercial, Uninsured)
•Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other
•Zip code, state and county or equivalent area of parent/caregiver's residence. Record FIPS if available

Stratification variables details
•Race/Ethnicity: Hispanic, Non-Hispanic Black, Non-Hispanic White; Non- Hispanic Asian/Pacific Islander, other Non-Hispanic
•Public vs Commercial (Private Insurance).
•HMO vs PPO vs FFS vs PCCM vs other; Within Medicaid, States may ask for reporting of FFS vs Managed Care or other relevant enrollment categories.
•Urban Influence Code. Identify the Urban Influence Code or UIC. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban- influence-codes.aspx#.UZUvG2cVoj8). Use parent or primary caregiver's place of residence to determine UIC. State and county names can be linked or looked up directly or zip codes can be linked to county indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to county or county equivalents as used in various states.

•Identify the Level of Poverty in the parent or primary caregiver's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using parent or primary caregiver's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:
o Lowest Quartile of Poverty if percent in poverty is <=12.5%
o Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
o Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
o First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
o Second Upper Quartile (>90th percentile)

Note: if needed, the Missouri Census Data Center may be used to link zip codes to county equivalents. http://mcdc.missouri.edu/

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Stratification by risk category/subgroup
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The rate should be reported stratified by age and within age strata stratified and by each of the stratification variables. Additional cross tabulation may be requested by the accountability entity. Biological risk for asthma ED use has not been shown to be associated with the specified sub-stratifying variables, but social determinants of health are associated with asthma care and utilization. Therefore we specify the measure to be reported as BOTH a single value for each age group and stratified by key covariates (e.g. race/ethnicity, insurance status, urbanicity, and poverty of county of residence).

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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

S.18. **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.)*
Step 1: Select starting cohort
Identify the upper age limit to be used, either 18 or 21. The measure is specified from 2 to 21 years, with 19-21 year olds considered optional at the discretion of the accountability entity.

Appendix Figures 1 and 2 and Appendix Table 1 provide an overview and guide for eligibility and sample selection.

Step 2: Conduct analysis of administrative data using the specifications described in denominator description to identify children within the specified age range with identifiable asthma. The analysis should be conducted on a month by month basis as described herein:

Determine eligibility for each patient, as of the last day of the month prior to the reporting month. For example, if the goal is to report for January 2011, first identify children with identifiable asthma (above), and analyze all of calendar year 2010 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2010. Next, for February analyze all of calendar year 2010 AND January 2011. Continuous enrollment criterion requires that the child was enrolled in December 2010 and January 2011. Repeat this progression monthly so that for December, one would identify children with identifiable asthma and analyze all of calendar year 2010 AND January through November 2011 when doing so. Continuous enrollment criterion requires that for December the child was also enrolled in October 2011 and November 2011. Appendix Figure A.1.a describes and illustrates the month by month analysis.

Step 3: Identify ED Visits and hospitalizations for asthma in eligible children.
Considering only the children who were identified as eligible in the given month
according to Step 2, perform a month-by-month analysis to identify and log all ED visits with asthma as a primary or secondary diagnosis and all hospitalizations with asthma as a primary or secondary diagnosis for each reporting month, using specifications described in denominator and the codes described above and in table 1 of the Appendix. Maintain stratification data elements, age, and unique identifiers.

Step 4: Stratify by age and develop random samples.
Stratify by age group (use age at month of qualifying event):
- Age 2-5 years (second birthday to the day before the 6th birthday);
- Age 6-11 years (sixth birthday to the day before the 12th birthday);
- Age 12-18 years (twelfth birthday to the day before the 18th birthday); and
- Age 19-21 years (nineteenth birthday to the day before the 21st birthday).

For each age group develop a random sample of 500 events as described in the sampling section below and illustrated in Appendix Figure 2.

Appendix Figure 2 is necessary to guide sample development. Several key remarks may help Figure 2 to be more understandable:

Before sample selection can be randomized, eligibility needs to be determined based on 3 key factors:

- Identifiable asthma diagnosis AND
- Month by month time analysis AND
- Asthma emergency department (ED) visit OR Asthma hospitalization

After eligibility is determined, the randomized sample can fall into one of three groups only:

A. Asthma ED visit only OR
B. Asthma hospitalization on same day as ED visit OR
C. Asthma hospitalization only

A. Asthma ED visit only qualifies for (at least) denominator inclusion
B. Asthma hospitalization on same day as ED visit qualifies for denominator AND numerator inclusion
C. Asthma hospitalization only needs further investigation to determine denominator inclusion

- Do NOT include in denominator if sample was not hospitalized from an asthma ED visit OR
- Do NOT include if ED visit was already in the sample under any criteria AND
- Remaining: Do include in Denominator AND Numerator

Step 5: Collect stratification data elements from administrative data.
Collect the following data elements for all eligible children in each randomized sample. These data elements are used for reporting stratified results. Entities that are interested in assuring large samples for specific stratified analyses may choose to incorporate a further stratified sampling scheme and oversample to assure that there is a sample size of 100-500 per stratification category (e.g. race or ethnicity of interest). Such a sampling scheme must employ an appropriate weighting system (using the reciprocal of the likelihood for selection as a weight, c.f. Rao, P., 2000. Sampling Methodologies with Applications. New York: Chapman & Hall) to estimate overall performance. Alternatively, the stratified samples may be used only for reporting stratum specific performance comparison and not for estimating the overall performance. Approximate 95% confidence interval widths (assuming a rate of 50% appropriateness) are shown in the sampling specifications. We specify to oversample by 25% to account for potential loss in our event identifications.

Stratification data elements include:

- Race
- Ethnicity
- Insurance type (Public, Commercial, Uninsured)
- Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other
- Zip code, state and county or equivalent area of parent/caregiver's residence. Record FIPS if available

Step 6: Categorize stratification variables as described in the stratification section S.12.

Step 7. Conduct Chart Audit (Medical Record Review) of GROUP A ED Visits.
Group A ED visits that have been selected for inclusion in the sample require a chart audit to assess eligibility for the numerator based on the explicit appropriateness criteria. They have already qualified for inclusion in the denominator. Eligibility for the numerator is established based on documentation of any of the following items. Review may be terminated once any qualification for the numerator is identified.

- Disposition of the child from the ED was to an inpatient hospital.
- Documented physical findings consistent with respiratory distress, including:
  - o Labored breathing with retractions and/or grunting; or
  - o Labored breathing with evidence of accessory muscle use; or,
  - o Markedly decreased breath sounds;
- Low O2 saturation level, defined as < 90%;
- An ABG obtained and reported;
- The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED;
- Specific documentation that:
  - o The child was referred to the ED after evaluation by the PCP or other licensed clinician practitioner; OR

. o The child received two or more doses of inhaled rescue medications without sufficient clinical improvement; OR
. o The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined "red zone" of peak flow measurement as part of a pre-specified asthma action or similar plan.

There is no specified order for review. Some institutions may prefer to record all reasons for numerator qualification to support ongoing or planned improvement activities.

Note 1: Evidence for hospitalization above requires that the child was admitted to any hospital as an inpatient. This includes admission directly to a medical or pediatric ICU or inpatient floor or transfer directly to an inpatient facility. If a child is transferred to another hospital, confirmation that the child actually was admitted directly (i.e., was not first admitted to another ED prior to admission) is necessary prior to qualifying for the numerator. Such confirmation may include evidence from the administrative data review in Step 2. Other potential sources for this information include ED discharge summary, disposition on a flow, admit, or discharge form, or documentation by doctors, nurses, nurse practitioners or physician assistants.

Note 2: Evidence that the child was referred to the ED requires documentation of both of two requirements. The requirements are:
- The child/adolescent was referred by a clinician to come to the ED; and
- The child/adolescent was evaluated by the clinician prior to referral. Generally such evaluations will be in person. Assessment of respiratory distress by listening or speaking to the child/adolescent over the telephone is sufficient if such an examination is clearly documented. Report of this requirement being met by the child/adolescent or parent/caregiver is sufficient to meet this criterion. Report of contact from the referring physician can also fulfill this criterion. Nursing notes, triage notes and clinician notes, particularly history of present illness (HPI) are common sources for this data.

Note 3: Evidence of a parent or caregiver report that the child received two or more doses of an inhaled rescue medication with insufficient clinical improvement typically will be found in triage, nursing, clinician, or respiratory therapy notes. It may also be documented as a part of medication reconciliation during intake. It requires documentation:
- That multiple treatments of medication were provided by inhalation or injection prior to arrival in the ED;
- That the medication(s) provided were specifically rescue medications and are not a part of the of the child/adolescent's preventive or maintenance regimen; and,
- That the child continued to be in distress following the treatments (alternately that the child did not improve substantially).

Note 4: Parent / caregiver report that their child was in a pre-defined "red zone" of peak flow measurement includes documentation:
- That a pre-specified asthma plan (action plan) exists and defines a "red zone" based upon an objective respiratory measurement, such as a peak flow rate; and
- That the objective assessment was made prior to coming to the ED and that the results were in the pre-specified "red zone."

Note 5: Reports of the physical exam typically may be found on triage, nursing, physician, nurse practitioner, physician assistant, or respiratory therapist notes. Diverse language may be used to describe similar findings, for example:
- The term pulling may be used to describe retractions. Retractions may be described as nasal flaring (particularly in infants), or by location (see below);
- Increased work of breathing may be indicated or it may be described by physical findings such as the use of accessory muscles, such as sub or intercostal muscles, supraclavicular or suprasternal. "Mildly" increased work of breathing or "minimal" retractions do not meet these criteria.
- Labored breathing, significant increased work of breathing, respiratory distress (moderate or greater), difficulty breathing, poor air entry (or air exchange or air movement) may all describe findings that meet this criterion. Grunting indicates that the child or adolescent is generating clearly audible sounds with each breath concomitant with apparent increased work of breathing. These may be found in the general description or respiratory section of the physical exam.
- Markedly (or severely) reduced breath sounds and descriptions of poor air movement are typically a part of an auscultation during the pulmonary exam.

Note 6: Documented evidence of the percent oxygen (O2) saturation from a transcutaneous assessment can be located in a flow sheet, nursing, respiratory therapy, or physician/nurse practitioner/physical assistant note or may be recorded as part of the physical exam. The O2 saturation may be obtained initially at triage and is often assessed periodically during the visit. Any O2 saturation less than 90 satisfies the criteria.

Note 7: An ABG requires drawing of a blood specimen from an artery and is distinguished from a venous blood gas, which would not fulfill this criterion. This typically would be found in a laboratory results section of the record or commented as a finding in a clinician's note, such as a respiratory therapist, doctor, PA, NP, or RN. An ABG is typically comprised of at least a pO2, pCO2, and pH.

Note 8: Consultation with a pulmonary specialist or other asthma specialist requires both an order for such a physician consultation and evidence that the consultation occurred, including a note from the consultant specialist. Typically a consultation from a pulmonologist, pediatric pulmonologist, allergist, or pediatric allergist would fill this criterion.

Identify which ED visits meet at least one criterion for the Numerator.
Maintain stratification variables.

Step 8: Conduct Chart Audit (Medical Record Review) to Assess Eligibility of GROUP C Hospitalizations for Inclusion in Denominator. Within each stratification group (as determined above), identify the asthma hospitalizations for which there were not associated ED visits (Group C) found in the administrative data. An asthma ED visit and asthma hospitalization are said to be associated on the basis of the administrative data review only if they occur on the same service data and at the same institutions and if the hospital discharge date is after the ED service date. Such hospitalizations should have been included in Group B. Other hospitalizations require a review of the medical record to determine if they were admitted or transferred directly from an ED visit that was not otherwise in the sample (i.e., was not identified via the administrative data analysis).

The chart audit/medical record review seeks evidence that the child was admitted to the hospital directly from the ED or transferred directly from another hospital's ED. Evidence may include an ED note (physician, nurse, physician assistant, nurse practitioner), flow, or face sheet that indicates the disposition of the ED visit was hospital admission.

It may also include a note from within the hospitalization (including the admission note or any physician, nurse, physician assistant, nurse practitioner note), flow sheet, face sheet, or discharge summary that indicates that the hospitalization came directly from (was admitted from or transferred directly from) an ED. In either case, the ED visit is only eligible for inclusion if the chart review specifies the date and institution of the ED visit sufficiently to assure that it can be uniquely identified and all duplication avoided. Others are excluded.

For example if an ED visit was identified in Group A and the resulting hospitalization appeared in Group C (either because of a different service date or different institution), the Group A ED visit would be included and the Group C hospitalization excluded as a duplicate (even though there was a preceding ED visit). If the child is uniquely included in the sample for that month and there is clear evidence that the admission came directly from an ED (e.g., was not transferred from another hospital after having been admitted from the ED) this measure can be satisfied. De-duplication requires the elimination of any duplications that remain in the sample, considering the unit of analysis to be the ED visit. In other words, all ED visits must be included only once. Further, an ED visit identified via the hospitalization that also was a transfer from another ED visit already in the sample should have been removed as a duplicate. Similarly all hospitalizations lacking sufficient document that the child was admitted or transferred directly from an ED visit or lacking sufficient detail to allow confirmation that the ED visit referred to in the notes is not already in the sample elsewhere (e.g., from Group A) should have been removed.

Those Group C hospitalizations that can be identified as resulting from a unique (unduplicated) ED visit are included in BOTH the numerator and the denominator.

Step 9: Calculate and report the measure.
a) For each age stratum, count the number of events in the sample that qualify for the denominator (ND).
b) For each age stratum, count the number of events in the sample and in the denominator that qualify for the numerator (NN).
c) For each stratum, calculate the percent of appropriate ED visits as Percent Appropriate = 100 * (NN / ND). Report to one decimal place.

Step 10: Report each stratification category listed below, that have an N of at least 50.
a) Race and ethnicity
b) Insurance type (Public/Medicaid, Private/Commercial, None, other)
c) Benefit type: HMO vs PPO vs FFS vs PCCM vs other
d) Urban Influence Code or UIC.

e) Level of poverty in the county of residence.

Step 11. Calculate and report 95% confidence intervals (using binomial distribution for each stratum) for each age specific stratum and for all of the Step 9 stratifications.

a) Calculate the standard error as the square root of each proportion by [1-the same proportion] divided by the number in the denominator.
b) Multiply the standard error by 1.96.
c) Subtract that value from the measured proportion. Report the greater of 0 and that number as the lower bound of the 95% confidence interval.
d) Add the product from b to the measured proportion. Use the lesser of that sum or 1 as the upper bound of the 95% confidence interval.
e) To report as percent, multiply by 100.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Within each age group, randomly select 500 ED visits among those identified in Step 4. Analyze each age strata's random sample distinctly:

Sort into three groups according to Appendix Figure A.1.b.

• Group A: Those with asthma ED visits ONLY and no associated asthma hospitalization to the same hospital on the same date. These ED visits are INCLUDED in the Denominator and receive Medical Record Review to assess eligibility for the Numerator;

• Group B: Those with both asthma ED Visits and asthma Hospitalizations at the same facility on the same date and for whom the hospital discharge date is after the ED date of service. These ED visits are INCLUDED in both the Denominator and in the Numerator. No further review is necessary to establish appropriateness;

• Group C: Those with asthma Hospitalizations ONLY and no associated asthma ED Visit to the same hospital on the same date. Please note that children admitted to the ED one date and admitted to the hospital the next day (from the same ED visit) will be identified in this group. Group C Hospitalizations are subject to Medical Record Review to assess eligibility for the Denominator. If they are eligible for the denominator they will be included in BOTH the Numerator and Denominator.

Please note that the terms medical chart and medical record are used interchangeably, as are the terms audit and review in this context.

Notes:
•Determining eligibility for sample selection precedes determining eligibility for measure.
•On the basis of the Administrative Data Analysis, children who are potentially eligible for the measure will be identified and segregated into Groups A, B, and C (the blue boxes above).
•Children are eligible for Group B if three things are found in the administrative data: ED Visit; Hospitalization on same day and same institution; and Hospital discharge is after date of ED visit.
•National and NY State data suggest that approximately ¾ of childhood asthma hospitalizations are admitted from ED, that about 1 in 9 childhood asthma ED visits result in hospitalization and that children admitted from the ED may not have their ED visit coded in administrative data.
•Medical record review determines eligibility for numerator among the Group A children, all of whom have already qualified to be included in the denominator.
•Group B children are eligible for both the numerator and the denominator on the basis of administrative data analysis alone and do not require chart review.
•Medical record review determines eligibility for inclusion in the measure (denominator!) for Group C children. If they are eligible for the denominator (i.e. that have been admitted directly from an unduplicated ED visit) then they are also qualified for the

numerator.

The impact of sample size on the width of the confidence interval is illustrated by assuming 50% appropriateness and a variety of sample size to calculate the width of the confidence intervals around the estimate obtained above. Variations from 50% will bring down the size of the confidence interval.

N= 50, + / - 13%
N= 75, + / - 11%
N= 100, + / - 10%
N= 150, + / - 8%
N= 200, + / - 7%
N= 250, + / - 6%
N= 400, + / - 5%

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
N/A

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System, Population : Community, Population : County or City, Population : National, Population : Regional, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospital/Acute Care Facility, Other
If other: Emergency Department

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Testing_Asthma5_Round_2_vFinal.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma: A PQMP Measure

Date of Submission: 1/6/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.23*)**	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

New York State Medicaid claims data 2010 – 2012.

Our work builds off of work performed by our CAPQuaM partner and steering committee member, NCQA. For specific data reliability and signal to noise analyses, we incorporate by reference (and will present more selectively) NCQA data relevant to their submission for NQF –endorsed asthma related measures:

- Use of Appropriate Medications for People with Asthma (ASM) – 0036 (we understand this is no longer being maintained as of 2015, but it was endorsed and the data were accepted.)
- Medication Management for People With Asthma (MMA) – 1799
- Asthma Medication Ratio (AMR) – 1800

We note that 1799 and 1800 are not directly applicable because they were tested at the score level. However, the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to produce valid scores.

The analyses above provide information regarding the capacity to use administrative data to identify the applicable denominator population. There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures.

The HEDIS score level testing alluded to above includes the following: To provide additional reliability evidence, NCQA conducted a field test, and retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims. The ultimate objective of the field test was to determine the ability of

health plans to reliably report complex administrative measures requiring multiple sources of data in addition to determining the completeness of the data for this specific population. Reliability was estimated on the HEDIS 2011 submissions (2010 data) using the beta-binomial model. Reliability used here is the ratio of signal to noise. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good. Reliability statistic from the HEDIS 2010 data set was calculated as 0.87141 for commercial and 0.93853 for Medicaid of for the overall ASM measure (0036), demonstrating that the data elements upon which the measure relies are reliable even in the context of a complex measure that is built upon them. We incorporated these findings into the design and specification of our measure.

We cite these not as specific evidence of score level performance of our measure, but as evidence that measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal to noise at a very high level.

Newly abstracted data was also used for this measure.

1.3. What are the dates of the data used in testing? 10/2009 – 11/2013

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

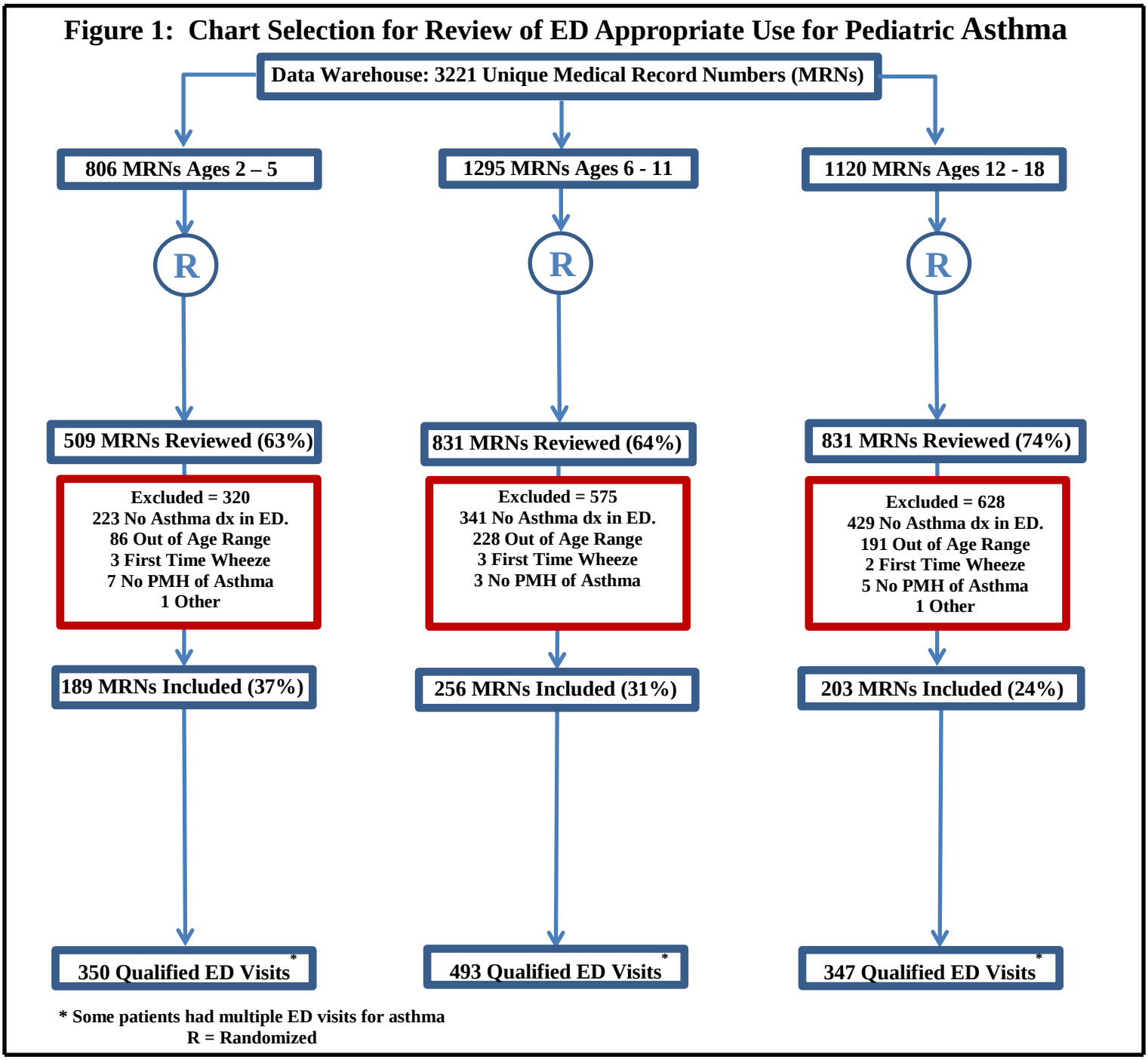
Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☒ health plan	☒ health plan
☒ other: Integrated delivery system, population, state, region, county	☒ other: Integrated delivery system, population, state, region, county

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Foundational analyses for this measure included:

Analysis of NY State Medicaid Managed Care claims data, including claims from all MCO's that are contracted for Medicaid care by our partner, the NY State Dept of Health. We identified eligible populations and events from both RY 2011 and 2012 and include children from counties in nine urban influence codes and in counties poverty level 1-3. NY State does not have any counties in the lowest 25% of poverty or with UIC of 10-12. New York has more than 60 counties and numerous health plan vendors. Analysis in year 2011 provided very similar data to 2012.

Figure 1: Chart Selection for Review of ED Appropriate Use for Pediatric Asthma



For the NCQA analysis, nine health plans covering a variety of geographic areas within the United States were asked to provide a complete administrative data file consisting of any member in their commercial and Medicaid product lines for anyone that had a diagnosis code for asthma during the calendar years of 2009-2010.

The complete member-level administrative file used for analysis included a total of more than 82,000 health plan members with asthma.

The specific measure demonstration and testing was done at one site, a New York City Academic Medical Center. In this testing, **sample selection can be summarized in the diagram above**.

The eligible observation period was October 2009 to November 2013. Please note, because of the limitations of the data systems available for testing randomization happened at the level of the patient. For patients with 1-3 visits for asthma in the included time frame, we included all visits. For patients with more than 3 visits, the first three visits were included. Hence the average number of visits per child was 350/189= 1.9 for the younger children, 493/256=1.9 for the school age children, and 347/203=1.7 for adolescents. In NY State Medicaid in 2011, the median number of visits per child was 1, the 75th percentile was 2 and the 90th percentile was 3 (N=26,169 children). Hence this finding is plausible and consistent, given the 4 year time frame that we sampled.

The eligible observation period was October 2009 to November 2013. Please note, because of the limitations of the data systems available for testing randomization happened at the level of the patient. For patients with 1-3 visits for asthma in the included time frame, we included all visits. For patients with more than 3 visits, the first three visits were included. Hence the average number of visits per child was 350/189= 1.9 for the younger children, 493/256=1.9 for the school age children, and 347/203=1.7 for adolescents. In NY State Medicaid in 2011, the median number of visits per child was 1, the 75th percentile was 2 and the 90th percentile was 3 (N=26,169 children). Hence this finding is plausible and consistent, given the 4 year time frame that we sampled.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Refer to Figure 1 in 1.5. Using the institutional data warehouse, we randomly identified medical record numbers of children who had both ED visits and asthma diagnoses in the specified time frame. Because we were not using claims data to select them, charts had to be reviewed for evidence of prior asthma and to assure that ED visit and asthma diagnosis were concurrent and in the selection time frame. ED visits were excluded if there was not evidence that they were known to be asthmatic, if the ED visit did not have asthma as the first or second diagnosis, or if the ED visit was not in the specified time frame. We included up to 3 visits per selected child, using the first 3 visits when more than three were present. Inclusion criteria included an ED visit with previously established asthma as a primary or secondary diagnosis as documented in the electronic medical record. We developed 3 samples stratified by age: 2-5 years, 6-11 years, and 12-18 years. For ages 2 – 5, we included 350 visits; ages 6 – 11, 493 visits; ages 12 – 18, 347 visits. So included in the measure testing was a total of 1200 ED visits were included in the chart review testing.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Assessment of the capacity to identify the eligible population and qualifying events was performed in NY State Medicaid data in both 2011 and 2012 reporting years.

Our construct for the CAPQuaM measure was defined by the multidisciplinary national expert panel using a RAND type modified Delphi process. The panel initially used the term persistent asthma to describe asthma that was pre-existing and should have been recognized as asthma by the health care system prior to the timing of the ED visit. This construct was renamed by our stakeholder group to be identifiable asthma to avoid

confusion with other uses of the term persistent asthma. The construct was intended to be more inclusive than HEDIS' persistent asthma diagnosis, while still removing from consideration those whose asthma was unlikely to have been actively managed at the time.

Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1% with the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. This ratio is 2.8, which is between 2-3, which is what we had predicted (based on the team's reading of the literature) and was the goals we were hoping to achieve with our criteria and was interpreted to suggest construct validity for our measure. Using data form the National Survey of Children's Health, we estimated the expected rate of asthma in the NY State Medicaid child population to be between 15 - 16%, indicating that our criteria did provide a meaningful filter as we had intended.

We found that by reducing the continuous enrollment period down to three months as was suggested by members of our steering committee that we could increases the number of children eligible for the measure by several tens of thousands while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group.

Assessment of data elements for identifying a population with asthma was performed by NCQA in nine geographically diverse managed care plans.

Assessment of appropriateness was performed in 1200 pediatric ED visits from a single medical center.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Insurance status and race/ethnicity for the single site analysis.

Race, ethnicity, zip code, level of poverty in the zip code of caregiver residence, and urban influence in the county of caregiver residence for the NY State analysis.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Validity testing was performed at the data element level for both the numerator and the denominator.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

_____

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Please see descriptions of both NCQA and CAPQuaM testing above in 1.2-1.7.

As described in the next section the literature also supports the use of claims data to identify the presence of asthma.

We develop our measure using scientifically sound principles. We first discuss research involving the soundness of our data sources, which include both administrative data to identify cases (and a fraction of numerator qualifications) and chart review (medical record audit) to confirm some denominator inclusions and to identify most numerator inclusion. This is a generally accepted and standard approach with acceptable reliability.

We use administrative data to identify the age of the child, various stratification variables and the presence of asthma, as well as the presence of an asthma ED visit or hospitalization. These are routinely used to support billing by CMS, Medicaid, and private insurers and are routinely used in quality measurement. Administrative data are not typically sufficient for detailed clinical assessment.[1-5] HEDIS developed a hybrid approach, using administrative data and chart review that this measure borrows heavily from. [6, 7]

There is moderate agreement (kappa = 0.45 – 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data. The agreement improves from 0.55 to 0.60 when using two years of data. (8). We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

The explicit criteria that we use were developed using a slightly modified version of the RAND/UCLA Appropriateness Method that maintained the key aspects of that approach, including a detailed literature review, a multidisciplinary and geographically diverse expert panel comprised of both clinicians and researchers, and the two Round modified Delphi Process. The general reliability of this approach is well established. [9, 10] It has been applied successfully to pediatric services previously. [11-13] We have used as criteria for this measure those specifications whose median rating is 8 or 9, the two highest ratings.

In our testing of the criteria during chart audit used a paper data collection instrument that was largely a checklist of yes/no for the various items. After a brief training by the physician who organized the testing three non-clinical research assistants (one MPH, 2 Bachelors) conducted chart audits. Kappa is presented in the next section.

1. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997.

35(6): p. 539-52.
2. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p. 199-207.
3. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p. 813-25.
4. Angier, H., et al., *Variation in outcomes of quality measurement by data source.* Pediatrics, 2014. **133**(6): p. e1676-82.
5. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.
6. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.
7. NCQA. *National Committee for Quality Assurance*. [cited 2014 7/30/14]; Available from: http://www.ncqa.org/
8. Huzel, L, et al. Diagnosing Asthma: The fit between survey and administrative database. Can Respir J. 2002 Nov-Dec;9(6):407-12.
9. Fitch, K., et al., *The RAND/UCLA Appropriateness Method User's Manual*. 2001 RAND.
10. Kosecoff, J., et al., *The appropriateness of using a medical procedure. Is information in the medical record valid?* Med Care, 1987. **25**(3): p. 196-201.
11. Kleinman, L.C., et al., *The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States.* Jama, 1994. **271**(16): p. 1250-5.
12. Kleinman, L.C., E.A. Boyd, and J.C. Heritage, *Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians.* Jama, 1997. **278**(6): p. 497-501.
13. Keyhani, S., et al., *Overuse of tympanostomy tubes in New York metropolitan area: evidence from five hospital cohort.* Bmj, 2008. **337**: p. a1607.

Table 1. 360 Degree Pediatric Quality Measure Development: Overview

Stage	Phase	Innovation	Product(s)
1. Clinical Criteria Development	a. Input Development	1. Focus groups of caregivers of children with asthma who have used the ED 2. Interviews with front line clinicians: primary care, asthma docs, and ED docs	1. Literature review 2. Summary of consumer perspectives, values and understanding relevant to clinical issue of interest 3. Summary of findings form clinician interviews
	b. RAND/UCLA 2 Round Modified Delphi Process	1. Inclusion of consumer perspectives as a key input; 2. Use of this method to identify appropriateness criteria in national performance measure development;	1. Explicit criteria that rank a comprehensive and mutually exclusive set of clinically detailed scenarios;
2. Boundary Guideline Development	Criteria Enhancement	1. Iterative process to enhance reliability and internal consistency of the explicit criteria set with a goal of outlining three boundary spaces	1. Internally consistent set of explicit criteria that are stable in their representation of the expert panel perspective. "Enhanced criteria"
	Guideline Articulation	1. Stakeholder (including experts, users, clinicians, consumers and others) informed review of the enhanced criteria. 2. Definition of zones of potential overuse, potential underuse, and professional interaction and decision-making based upon the explicit criteria 3. Stakeholder valuations of potential deviations from guideline 4. Boundary Guideline	1. Boundary Guideline 2. Prioritization list

3. Creation of Measure	Specification	1. Translation of guideline into specification of necessary data 2. Iterative process to define optimally efficient sources of data to allow for measurement and stratification	1. Initial specification of measure
	Review	1. Constructive peer review of specifications by stakeholders in Steering Committee and SAB	1. Final specifications of measure including variables for stratification as needed
	Fielding and testing of measure	1. Measure testing	1. Functional experience and practical understanding of measure, its scoring, variability, and interpretation

This measure was developed and assessed using a pre-specified process and consistent with CAPQuaM's peer reviewed 360 degree method outlined in the table above.
Explicit criteria were developed using a variation of the two-round modified Delphi process RAND/UCLA Appropriateness Method with a multidisciplinary and geographically diverse expert panel comprised of both clinicians and researchers. Identifiable asthma was based on panel findings and appropriateness criteria included for this measure were those that were both available in the chart and highly rated.

Development included a series of alpha tests to refine specifications by conducting iterative analyses in New York State Medicaid data. Conclusions from alpha tests include:

1) The reporting period and the assessment period could not overlap completely, leading to use of 2 years of data as shown in the specifications' diagram. The optimal approach was to divide the reporting year into 12 reporting months. ED events in that month are eligible for the numerator if persistent asthma criteria have been satisfied (combining the look-back year and all prior months in the reporting year) and the child has been continuously enrolled for the two months immediately prior to the reporting month. The optimal unit for the denominator is in child-months;
2) Using both revenue codes and CPT codes increased our sensitivity meaningfully, a choice validated by consultation with coding and billing experts;
3) NY State Medicaid data and national survey data (HCUP) converged to demonstrate the importance of including hospitalizations as numerator events even when the underlying construct is ED visits. This is consistent with policies of many payers to request providers not to submit both ED and hospital claims for the same day. Error would be far less by considering both ED visits and hospitalizations as numerator events, than by not including hospitalizations.
4) The expert panel only wanted numerator events for which the children were already known to the accountable entity as having asthma and established definitions for such "identifiable asthma". Identifiable asthma was intended to be more restrictive than the 15-16% identified by our analysis of the 2011 NSCH as having ever been told they had asthma and much less restrictive than the HEDIS definition of persistent asthma. Alpha testing in NY State Medicaid demonstrated the expected results:
    a. Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1%, the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. This ratio is 2.8 (our predicted and target result was between 2-3 based the literature achieve and our intended construct).
    b. Relaxing the continuous enrollment period to 3 months was suggested by members of our stakeholder steering committee. Doing so increased the eligible number by several tens of thousands while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children. [1] Use of the medical record as a valid source of information to judge appropriateness is well accepted. [2] Chart audits are used frequently to generate research in Emergency Medicine. [3, 4]

Key panel ratings are shown. Constructs rated 7 or higher are endorsed, 8 or higher strongly endorsed, and 2 or lower strongly rejected.

Scenario	**MED**
Wheezing on presentation to the ED establishes that the ED was an appropriate level of care for that child.	5
Retractions or labored breathing during the ED visit establishes that the ED was an appropriate level of care for that child.	**9**
Decreased breath sounds establish that the ED was an appropriate level of care.	6
Markedly decreased breath sounds establish that the ED was an appropriate level of care.	**7**
Obtaining an ABG in the ED establishes the ED as an appropriate level of care for that child.	**9**
Oxygen saturation less than 90% establishes that the ED was an appropriate level of care for that child.	**9**
Hospitalization following the ED visit establishes that the ED was an appropriate level of care for that child.	**9**
An ED visit less than 72 hours following a previous ED visit in a child with asthma establishes that the ED was an appropriate level of care for that child.	4
Prescription of an oral steroid burst establishes that the ED was an appropriate level of care for that child.	4
An ED visit less than one week following a hospital discharge in a child with asthma establishes that the ED was an appropriate level of care for that child.	4
An ED visit less than 72 hours following a hospital discharge in a child with asthma establishes that the ED was an appropriate level of care for that child.	3
A specialty consultation in the ED establishes that the ED was an appropriate level of care for that child.	**8**
Homelessness establishes that the ED was an appropriate level of care for that child.	3
Parent report that the PCP is generally unavailable for urgent asthma care establishes that the ED was an appropriate level of care for that child.	5
Parent report of inability to reach the PCP during the current event establishes that the ED was an appropriate level of care for that child.	6
Parent report that they were referred into the ED by phone contact a clinician establishes that the ED was an appropriate level of care for that child.	**8**
Parent report that they were referred to the ED after being seen by a clinician establishes that the ED was an appropriate level of care for that child.	**9**
Parent report that the child did not respond to a dose of a rescue medication establishes that the ED was an appropriate level of care for that child.	6
Parent report that they are unable to afford needed asthma medications establishes that the ED was an appropriate level of care for that child.	3
Parent report that they are unable to obtain needed care because of financial barriers establishes that the ED was an appropriate level of care for that child.	3

The proportion of visits found to be appropriate varied by age and there are biological reasons that make plausible such differences not only being related to health services. Therefore we have specified this measure to be reported as stratified by age. Our data showed that within the 2 – 5 year age group, 54.3% were appropriate, within the 6 – 11 year age group, 44.3% were appropriate and within the 12 – 18 year age group, 48.3% were appropriate, p =.019. The breakdown is as follows:

- For children 2-5: 181 of 335 audits (54.3%) were deemed appropriate.
- For children 6-11: 209 of 477 audits (43.8%) were deemed appropriate.
- Adolescents aged 12-18: 165 of 341 audits (48.4%) were deemed appropriate based upon information in the chart audit.

Criteria for appropriateness that were met were recorded and did vary by age.

1. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.
2. Kosecoff, J., et al., *The appropriateness of using a medical procedure: is information in the medical record valid?* Med Care,

1987: p. 196-201.
3. Gilbert, E.H., et al., *Chart reviews in emergency medicine research: where are the methods?* Ann Emerg Med, 1996. **27**(3): p. 305-308.
4. Worster, A., et al., *Reassessing the methods of medical record review studies in emergency medicine research.* Ann Emerg Med, 2005. **45**(4): p. 448-51.

The development team's goal was to develop an ICD10 code set that was fully consistent with the intent of the original measure. Our process began by performing general equivalency mapping using the forward mapping from www.icd9data.com. We then did a de novo review of the CMS ICD 10 CM set to seek to identify codes that might be appropriate for asthma. We reviewed potential codes identified by both sources and developed a new list of codes appropriate for inclusion criteria and a new list of codes appropriate for exclusion criteria. Drs. Kleinman and Sharma reviewed the lists independently and then achieved consensus in a conference call review and discussion. Key team members for this work were Suzanne Lo, MPH who staffed and coordinated this work, Sandeep Sharma, MD, Dr.PH and Lawrence Kleinman, MD, MPH. Dr. Sharma was a lead developer for one of CAPQuaM's 2 asthma measures and Dr. Kleinman is both CAPQuaM PI and was a lead developer for both measures. The guidance for the intended constructs for both ICD9 and ICD10 coding were the findings from a RAND style modified Delphi panel that incorporated 9 national experts over the course of the measure development process.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
IDENTIFYING A POPULATION WITH ASTHMA:

For the foundational NCQA work, NCQA's field test retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. Using the dataset provided, NCQA examined several different scenarios to determine the effects of different specification criteria on this particular population. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims.

Score level reliability of the HEDIS 2011 submissions (2010 data) was assessed using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

We cite these not as specific evidence of score level performance of the submitted measure, but as evidence that the HEDIS measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal to noise at a very high level. If the population assessment were inadequate, then these other measures which use the same data elements to establish their denominators could not achieve such high reliability scores. This is because failure to distinguish signal from noise at the level of the HEDIS denominators would lead to non-differential misclassification error which is a major bias towards the null, in other towards noise and away from signal. Hence these provide strong indirect evidence of the validity of our approach to capturing the measure's denominator.

While there is moderate agreement (kappa = 0.45 - 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data to parent report, the agreement improves from 0.55 to 0.60 when using two years of data. (Huzel, L. et al. Diagnosing Asthma: The fit between survey and administrative database. Canada Resp. Journal 2002.) We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

Further, we identify asthma visits and medications using the same data that an insurance company or Medicaid would use for payment, including ICD9 codes, CPT codes, and revenue codes. We have had conversations with expert coders and New York State Department of Health Office of Health Insurance Programs to confirm our choices.

The literature also supports our work. ICD-09 and ICD-10 codes for asthma on patients' medical charts typically match claims data. ICD-9-CM administrative data have been validated using various methodologies for various purposes (5-17). Studies have shown high sensitivity and specificity for diagnoses obtained from administrative data among children with high-risk conditions including asthma, (18), and high predictive value among adolescents and adults with asthma. (19) (20) HEDIS criteria using administrative data support peer reviewed research, for example in patients with persistent asthma based on HEDIS criteria in five Medicaid programs (Colorado, Georgia, Indiana, New Jersey, Washington) using ICD-9-CM code 493.x (21). Fowles and colleagues report sensitivity and specificity of claims compared with ambulatory medical records to identify asthma was 0.82 and 0.99, respectively. (22) Wilchesky compared chart abstraction to diagnoses obtained from administrative database: asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0). (23) Bronstein et al found that 88.3% of diagnoses asthma on claims agreed with medical record, with a negative predictive value of 0.85 and a positive predictive value of 0.88.They conclude that claims are generally an accurate indicator of the content of a patient encounter. (24) Steinwachs et al. compared billed claims to medical records based on date of visit and diagnosis, they found for asthma there was 90.9 percent of billed visits in record on same date and 82.8 percent of billed visits with same diagnosis in record on same date. (25) Quan et al documented the validity of ICD-9-CM and ICD-10 coding systems in coding clinical information and found that ICD-10 data was generally comparable with that of ICD-9-CM data in recording clinical information. (26)

From a public health perspective, asthma surveillance systems in several states, including Maine, North Carolina, Connecticut and Michigan, have shown the feasibility of using administrative data to identify children having asthma, based on primary and secondary diagnosis codes reported on inpatient and outpatient claims. (27-30) Researchers also classified children with evidence of persistent asthma using HEDIS criteria, (31). Another study showed the usefulness of ICD9 493.x to identify asthma for a quality measure using Maryland data. Like our measure, those researchers excluded children with a diagnosis of cystic fibrosis (ICD9 277). (32) regarding our capacity to identify exclusions, Quan et al found that claims had a PPV of 91.9, and a negative predictive value of 92.6, with *k* of 0.65 (substantial agreement[1]) compared to chart review. FICD 10 performed similarly in this study.

REFERENCES

(1) National Committee for Quality Assurance. (1 994). Health care quality improvement studies in managed care set1ing.s: A guide for state medicaid agencies. Washington, DC: Author.

(2) Leatherman, S., Peterson, E., Heinen, L., & Quam, L. (1991). Quality screening and management using claims data in a managed care setting. Quality Review Bulletin, 17, 349-359.

(3) Parente, S.T., Weiner,J.P., Garnick, D.W., Richards, T.M., Fowles,J., Lawthen, A.G., Chandler, P., & Palmer, R.H. (1995). Developing a quality improvement database using health insurance data: A guided tour with applications to Medicare's National Claims History File. American Journal of Medical Quality, 10(4), 162-176.

(4) Weiner, J.P., Powe,. R., Steinwachs, D.M., & Dent, G. (1990). Applying insurance claims data to assess quality of care: A compilation of potential indicators. Quality Review Bulletin, 16, 424-438.

(5) Best, W.R., S.F. Khuri, M. Phelan, K. Hur, W.G. Henderson, J.G. Demakis, and J. Daley. 2002. ''Identifying Patient Preoperative Risk Factors and Postoperative Adverse Events in Administrative Databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program.'' Journal of American College of Surgeons 194: 257–66.

(6) Geraci, J.M., C.M. Ashton, D.H. Kuykendall, M.L. Johnson, and L. Wu. 1997. ''International Classification of Diseases, 9th Revision, Clinical Modification Codes in Discharge Abstracts are Poor Measures of Complication Occurrence in Medical Inpatients.'' Medical Care 35: 589–602.

(7) Hsia, D.C., C.A. Ahern, B.P. Ritchie, L.M. Moscoe, and W.M. Krushat. 1992. ''Medicare Reimbursement Accuracy under the Prospective Payment System, 1985 to 1988.'' Journal of the American Medical Association 268: 896–9.

(8) Iezzoni, L.I., S. Burnside, L. Sickles, M.A.Moskowitz, E. Sawitz, and P.A. Levine. 1988. ''Coding of Acute Myocardial Infarction. Clinical and Policy Implications.'' Annals of Internal Medicine 109: 745–51.

(9) Jollis, J.G., M. Ancukiewicz, E.R. DeLong, D.B. Pryor, L.H. Muhlbaier, and D.B. Mark. 1993. ''Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research.'' Annals of Internal Medicine 119: 844–50.

(10) Lee, D.S., L. Donovan, P.C. Austin, Y. Gong, P.P. Liu, J.L. Rouleau, and J.V. Tu. 2005. ''Comparison of Coding of Heart Failure and Comorbidities in Administrative and Clinical Data for Use in Outcomes Research.'' Medical Care 43: 182–8.
(11) Muhajarine, N., C. Mustard, L.L. Roos, T.K. Young, and D.E. Gelskey. 1997. ''Comparison of Survey and Physician Claims Data for Detecting Hypertension.'' Journal of Clinical Epidemiology 50: 711–8.
(12) Quan, H., G. Parson, and W. Ghali. 2002. ''Validity of Information on Comorbidity Derived from ICD-9-CM Administrative Data.'' Medical Care 40: 675–85.
(13) Romano, P.S., B.K. Chan, M.E. Schembri, and J.A. Rainwater. 2002. ''Can Administrative Data Be Used to Compare Postoperative Complication Rates across Hospitals?'' Medical Care 40: 856–67.
(14) Romano, P.S., and D.H. Mark. 1994. ''Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment.'' Medical Care 32: 81–90.
(15) Weingart, S.N., L.I. Iezzoni, R.B. Davis, R.H. Palmer, M. Cahalane, M.B. Hamel, K. Mukamal, R.S. Phillips, D.T. Davies Jr., and N.J. Banks. 2000. ''Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program.'' Medical Care 38: 796–806.
(16) Yasmeen, S., P.S. Romano, M.E. Schembri, J.M. Keyzer, and W.M. Gilbert. 2006. ''Accuracy of Obstetric Diagnoses and Procedures in Hospital Discharge Data.'' American Journal of Obstetrics and Gynecology 194: 992–1001.
(17) Quam, L., Ellis, L. B. M., Venus, P., Clouse, J., Taylor, C. G., & Leatherman, S. (1993). using claims data for epidemiologic research. the concordance of claims-based criteria with the medical recordand patient survey for identifying a hypertensive population. Medical Care, 31(6), 498-507.
(18) Daley MF, Barrow J, Pearson K, Crane LA, Gao D, Stevenson JM, et al. Identification and recall of children with chronic medical conditions for influenza vaccination. Pediatrics 2004;113(1 Pt 1):e26-33.
(19) Vollmer WM, O'Connor EA, Heumann M, Frazier EA, Breen V, Villnave J, Buist AS. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol 2004;57:392-7.
(20) Twiggs, J. E., Fifield, J., Apter, A. J., Jackson, E. A., & Cushman, R. A. (2002). Stratifying medical and pharmaceutical administrative claims as a method to identify pediatric asthma patients in a medicaid managed care organization. Journal of Clinical Epidemiology, 55(9), 938-944.
(21) Samnaliev, M., Baxter, J. D., & Clark, R. E. (2009). Comparative evaluation of two asthma care quality measures among medicaid beneficiaries. Chest, 135(5), 1193-1196.
(22) Fowles, J. B., Fowler, E. J., & Craft, C. (1998). validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. Journal of Ambulatory Care Management, 21(1), 24-34.
(23) Wilchesky, M., Tamblyna, R. M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. Journal of Clinical Epidemiology, 57, 131-141.
(24) Bronstein, J. M., Santer, L., & Johnson, V. (2000). The use of medicaid claims as a supplementary soure of information on quality of asthma care. Journal for Healthcare Quality, 22(6), 13-18.
(25) Steinwachs, D. M., Stuart, M. E., Scholle, S., Starfield, B., Fox, M. H., & Weiner, J. P. (1998). A comparison of ambulatory medicaid claims to medical records: A reliability assessment. American Journal of Medical Quality, 13(2), 63-69.
(26) Quan, H., Li, B., Saunders, L. D., Parsons, G. A., Nilsson, C. I., Alibhai, A., et al. (2008). Assessing validity of icd-9-cm and icd-10 administrative data in recording clinical conditions in a unique duallycoded database. HSR: Health Services Research, 43(4), 1424.
(27) Deprez RD, Asdigian NL, Oliver LC, Anderson N, Caldwell E, Baggott LA. Development of a prototype system for statewide asthma surveillance. Am J Public Health 2002;92:1946-51.
(28) Buescher PA, Jones-Vessey K. Using Medicaid data to estimate state- and county-level prevalence of asthma among low-income children. Matern Child Health J 1999;3:211-6.
(29) Children's Health Council. Asthma affects nearly one in ten children enrolled in HUSKY Part A in 1999–2000. Hartford (CT): Children's Health Council; 2001.
(30) Dombkowski, K. J., Wasilevich, E. A., & Lyon-Callo, S. K. (2005). Pediatric asthma surveillance: Using medicaid claims. Public Health Reports, 120, 515.
(31) National Committee for Quality Assurance. Use of appropriate medications for people with asthma. HEDIS 2003, volume 2: technical specifications. Washington: NCQA; 2003.
(32) Stuart, B., Singhal, P. K., Magder, L. S., & Zuckerman, I. H. (2003). How robust are health plan quality indicators to data loss? A monte carlo simulation study of pediatric asthma treatment. HSR: Health Services Research, 38(6), 1547.
(33) Schneeweiss, S., & Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. Journal of Clinical Epidemiology, 58(4), 323-337.

NUMERATOR DATA ELEMENT ASSESSMENT:

We assessed the reliability of data abstraction by making a total of 30 comparisons per construct with our gold standard reviewer. 6 constructs were specific reasons for appropriateness (like markedly reduced breath sounds, or evidence of respiratory distress) and the 7th was the visit level assessment of appropriateness. We measured at the beginning of data collection following training, and again at the conclusion of data collection.

Prior to training, we conducted 30 comparisons to the gold standard reviewer and calculated positive and negative predictive values, sensitivity and specificity, and kappa scores and found PPV ranged from 0.78-1, NPV from 0.71-1, Se from 0.60 – 1.0 and kappas were 0.44, 0.6-0.60. 0.67,0.79, and 1.0.

Aggregating the 180 comparisons, PPV=0.89, NPV=0.91, Se=0.76, Sp=0.96, and kappa=0.757, indicating very good to excellent agreement at the data element level. The single most important finding regards assessing appropriateness for each visit, for which there were 30 comparisons:

PPV	0.95
NPV	0.80
Sp =	0.888889
Se=	0.904762
Kappa	0.769231

indicating excellent reliability at the measure numerator level.

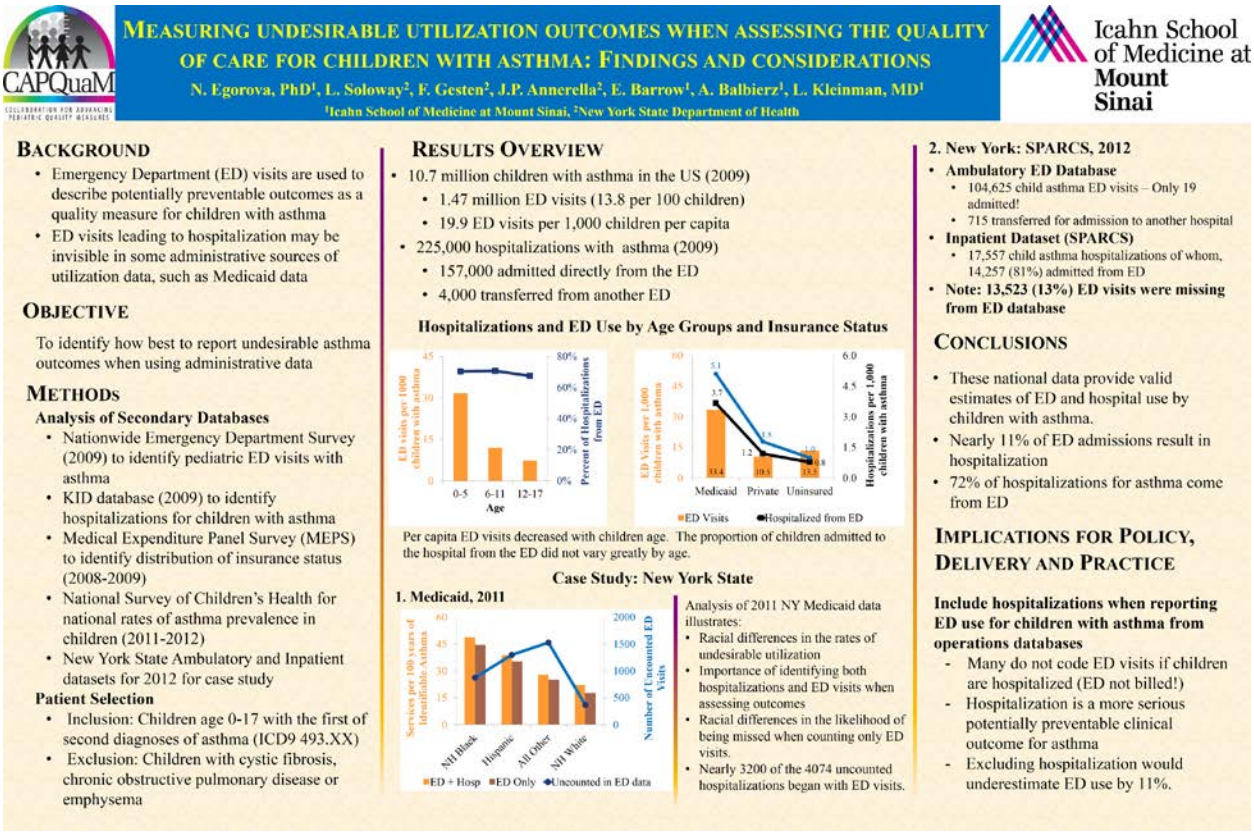
After experience reviewing charts we repeated this assessment and findings were comparable, with the details for the Appropriateness assessment shown here:

PPV	0.882353
NPV	1
sp	0.866667
se	1
Kappa	0.866667

We conclude that research assistants were able to be simply trained to reliably asses the appropriateness of ED visits using the explicit criteria developed by an expert panel and that their performance improved and did not deteriorate with experience.

Our own research looking at NY State Medicaid and national all payer data (see poster presented at peer-reviewed AcademyHealth national meeting) is consistent with expert and other recommendations that to identify all ED visits, one also needs to include hospitalizations for asthma as potential indicators of an otherwise unrecognized ED visit, which we have done and incorporated into the specifications.

This is the poster presenting our original research regarding the inclusion of hospitalizations when considering potential inclusion in the denominator. Final inclusion requires evidence of an ED visit.



2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

Our interpretation is that administrative data are reliable for identifying asthma, and that year to year test retest reliability seems to indicate similar patterns of performance when identifying ED visits for asthma, reinforcing the reliability of our operational definitions for identifying eligible children. Our specification provide a sensitive and face valid approach to identifying an unbiased sample of children with ED visits(ensuring we don't bias the results towards the inappropriate by missing those with hospitalization).

Most databases contain consistent elements, are available in a timely manner, provide information about large numbers of individuals, and are relatively inexpensive to obtain and use. Validity of many databases has been established, and their strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented (30). Administrative data are supported, if not encouraged by federal agencies, such as NIH, AHRQ, HCFA, and the VA. The Centers for Medicare & Medicaid Services has made clear to the participating AHRQ-CMS CHIPRA Centers of Excellence funded to develop measures in the Pediatric Quality Measures Program that it places a premium on feasibility when assessing those measures that it will most highly recommend to states to complete. The sources of data for the existing measure and other similar measures are typically based upon administrative data as well, providing consensual validation for using administrative data as the primary data source.

Our Kappa results indicated excellent agreement in the reliability of the chart audit. Kappa values over 0.75 are considered excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions were only included if they were endorsed by the expert panel. In studying the denominator we found that a very few percent of potentially eligible children (<=2.5%) were excluded by clinical diagnoses. The use of three months of continuous enrollment was recommended by our multi-stakeholder consortium and avoids the exclusion of more than 20% of otherwise eligible children from the population with identifiable asthma compared to a 12 month requirement.

Denominator Exclusions
Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 Code: 496), Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03, 277.09), or Emphysema diagnosis (ICD-9 code 492xx).

Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month, as well as the index reporting month itself.

There are no numerator exclusions.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions are clinical and represent construct validity rather than statistical considerations.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☒ **Stratification by** 1 **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Specifications for this measure requires stratification by age group. Several additional stratifications are optional but may be required by the accountability entity. These variables include race/ethnicity, rurality/urbanicity and county level of poverty.

Within age group, we specify a number of stratifications as we have done for all of our CAPQuaM PQMP measure. Absent clear biological evidence that ED visits should be more likely in any of the sub categories we have chosen not to adjust but to report both topline and stratified results.

The NIH NHLBI NAEPP (http://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/full-report) guideline notes that goals of care and definition of successful management are the same regardless of baseline presentation. Hence clinical risk adjustment is not appropriate.

(page 38)

> ***"An important point linking asthma severity, control, and responsiveness is that the goals are identical for all levels of baseline asthma severity.*** *A patient who has severe persistent asthma compared to a patient who has mild persistent asthma, or a patient who is less responsive to therapy may require more intensive intervention to achieve well-controlled asthma; however, the goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by therapeutic intervention."*

High levels of appropriateness suggest that the children in the ED are there because of an immediate clinical need and the ED service is well utilized. Some of these may have been preventable with better quality care prior to the ED visit and some will not. When appropriateness sis high, Asthma ED visit rates represent a strong proxy for asthma clinical outcomes.

Low levels of appropriateness suggest that the cause of many ED visits is not break through asthma or failures of biological asthma management, but insufficient access or quality of care provided that families are seeking care in the ED as preferential to a less acute setting. The good news in such a finding is that clinical asthma outcomes are better than would appear simply by counting the number of ED visits.

The results have independent meaning but from both accountability and improvement perspectives there is synergy in the interpretation of this measure with the CAPQuaM rate of ED visits in asthma measure.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

The conceptual model is that of CAPQuaM that includes that in pediatrics age is a key predictor and stratification is valuable. We were asked by AHRQ and CMS to include other constructs and we have manifest them as specified, such as race/ethnicity, poverty level in the caregivers county of residence, rurality/urbanicity on the caregiver's county of residence, insurance type and plan type, when variable. We have not added a stratum for children with special health care needs since asthmatics going to the emergency room are highly likely to belong in this category.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis*

was used)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

For results of age-stratified analysis, please refer to section 2b4.4a

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Our medical chart audit found that the measure varies by race/ethnicity.

Appropriateness varied by age (Chi^2=8.2,p=.02), with younger (p=.01) and school aged (p=.01) children each being significantly different; Adolescents experienced a level of appropriateness intermediate to the other two groups and were not significantly different from them when combined (ie comparing Adolescents to All others). We also found racial differences with Hispanics at 44.1% appropriateness, non-Hispanic Blacks at 51.3%, Whites at 56.5% and all others at 72.2%. Chi square with 3 degrees of freedom was 15.4, with p=.0015. The appropriateness of ED visits for Hispanic children was less than for other children (p=.002).

Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (46.8%), p=.002. Black children showed a trend toward more questionable use compared to all other children (53.6% questionable vs 48.7%, p=.10).

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Contingency table analysis with chi square, SAS 9.4 Generalized linear models (Proc GLM) and SAS 9.4 Logistic Regression (Proc Logistic) analyses were performed and were coherent and each illustrated the presence of statistical differences among identifiable subgroups.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

The proportion of visits found to be appropriate varied by age and there are biological reasons that make plausible such differences not only being related to health services. Therefore we have specified this measure to be reported as stratified by age. Our data showed that:

- For children 2-5: 181 of 335 audits (54.3%) were deemed appropriate.
- For children 6-11: 209 of 477 audits (43.8%) were deemed appropriate.
- Adolescents aged 12-18: 165 of 341 audits (48.4%) were deemed appropriate

based upon information in the chart audit.

Criteria for appropriateness that were met were recorded and did vary by age

The GLM models regressed appropriateness simultaneously on the class variables Age Group, Ethnicity, Gender, and presence or absence of private insurance found that gender (P=.017), Hispanic ethnicity (p=.002), and private insurance (p=.005) were all significantly associated with level of appropriateness, as was age group (p=.009). For this analysis, N=1,188 with a model F value of 6.56 (Pr>F is <0.0001).

To confirm the distinction between what we expected to be strong and weak effects, we substituted day of week for the various demographic variables other than age group. The P value for day of week (as a class variable) was >0.30. The non-zero effect size is consistent with social science literature that suggests that variables such as time of day and day of week are weakly meaningful. Still, the lack of a significant finding in a reasonably good-sized data set demonstrates that spurious significant findings are not likely to be identified as significant.

Differences between major subgroups were statistically significant, including race/ethnicity, age group, and insurance status. We note that this is a stricter test than had the measure been assessed across different entities. These data showed differences by type of insurance, which in this case can serve as proxy for health plan. The F Value for Insurance Status (after controlling for age group) was 3.91 with 4 degrees of freedom, which exceeded the critical value and is associated with a p-value <0.004. This should correlate with excellent capacity to distinguish between health plans.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

In sum, we found true signal in social determinants (consistent with the asthma literature) and did not incorrectly identify weak signal as meaningful. The measure distinguishes signal from noise.

The measures are sensitive enough to detect meaningful differences as observed within a population (as described above). Since the sum of squares across populations is expected to be greater in distinct populations, we expect the measure to perform very well when comparing across populations as well. Since the effective sample size of within population comparisons (such as we have conducted) is diminished by an (unmeasured) intraclass correlation coefficient, we would expect greater power for equal sample size to detect differences

between entities than we had in our testing of various subpopulations within a single state. This supports the same conclusion. The signal to noise ratio is very strong for this measure.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Chart review data has been shown to be an accurate method for identifying the presence or absence of conditions required to identify the level of appropriateness of a clinical service. Documentation is a part of the clinical responsibility and failure to document is a quality deficit that is not construed as missing data. Since inclusion requires the affirmative presence of data and we are unaware of any evidence to suggest that there would be differential absence of data between appropriate and non-appropriate visits we are not concerned about introducing bias in our findings. Further, we use random sampling of eligible visits as another means to avoid the introduction of bias. (1, 2)

1. Kahn, KL, Kosecoff, J, Chassin, MR, et al. Measuring the clinical appropriateness of a procedure: Can we do it? Medical Care, 1988, 26:415-422.
2. Kosecoff, J., et al., The appropriateness of using a medical procedure. Is information in the medical record valid? Med Care, 1987. **25**(3): p. 196-201.

While assessing the definition for identifiable asthma, our colleagues at NY State Medicaid conducted a series of iterative analysis using NY State Medicaid Managed care data to assess the importance of our data elements and definitions. These analyses helped to confirm the importance of using, for example, both revenue codes and procedure codes to identify ED visits. These analyses also confirmed that the use of pharmaceutical data to identify children with asthma expanded the pool of these so identified and quantified that statewide doing so added around 10,000 to a total of around 190,000 children with identifiable asthma in the state. We found no evidence that this was a threat to the measure's validity. The key reason for inclusion of pharmacy data is that our expert panel directed us to use it and it is a slightly more sensitive way to identify asthmatic children from the pool of all children with asthma related claims. The expert panel did not want the absence of pharmacy data to preclude inclusion of a reporting entity in the measure or to exempt any entity from measurement. We do not have either direct access to the data or a copy of all the iterative analyses at this time or we would include more specific data to demonstrate these findings. The analyses were in hand and were incorporated into our decision-making at the time that we developed the specifications.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
See section above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
Not biased.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**
Detailed clinical data are needed. There are no technical barriers to capturing the necessary data in defined electronic fields in electronic health records. We view NQF endorsement as a step to help us to initiate a conversation to consider such inclusion.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
We have learned that chart review is a reliable and accepted method of measuring appropriate use. There are no technical barriers to incorporating structured fields to help assess the appropriateness of the visits in conjunction with the criteria outlined above and implemented in this measure, although such fields do not currently exist. We further demonstrated that our measure was able to identify differences in the proportion appropriate, such as those associated with age and race. For example, the overall level of appropriateness for children aged 2-5 was 54%, for children aged 6-1 was 44%, and for adolescents between 12 and 18, 48%. Because of these differences we have chosen to present the measure stratified by age group. We found that use of a clinical database was an inefficient way to identify eligible charts and thus have adapted eligibility criteria that rely on administrative data. Because chart review is relatively time consuming, we have articulated the specifications in a way that represents a hybrid whereby administrative data can qualify a proportion of numerator events without chart review. Our paper data collection tool underwent a number of revisions for time and data collection efficiency and the chart review team demonstrated excellent agreement in data collection with a group kappa of .923 in identifying numerator events. Although the chart collection tool is not a formal part of this measure, we would be happy sharing a general version (data collection template) of it upon request.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None at present.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting	

Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

N/A

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Some potential users are awaiting NQF endorsement.

The topic of ED asthma overuse was assigned to our measure development project in the Pediatric Quality Measures Program by CMS, by far the largest single third party payer for medical care for children in the US, and by AHRQ. Major federal policy makers have indicated to us that these measures are a priority. This measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

We have begun a dialogue with the CDC to consider use of this measure to serve their interests.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The CAPQuaM team includes multiple stakeholders, including payers and state agencies. Several are interested in using this measure and are awaiting NQF endorsement. As a part of our CAPQuaM work we will disseminate and assist in the implementation of this measure subsequent to endorsement. This measure has been approved for inclusion in the National Quality Measures Clearinghouse.

As noted above, the topic of ED asthma overuse was assigned to our measure development project in the Pediatric Quality Measures Program by CMS, by far the largest single third party payer for medical care for children in the US, and by AHRQ. Major federal policy makers have indicated to us that these measures are a priority. This measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

We have begun a dialogue with the CDC to consider use of this measure to serve their interests.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**

- **Geographic area and number and percentage of accountable entities and patients included**

N/A

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

Not in longitudinal use. As noted above, both high and low levels of appropriateness are interpretable and actionable as outcomes of asthma management. This measure of process provides information regarding the outcomes of asthma care – both access to care and quality of management. Its interpretation is synergistic with the CAPQuaM rate of asthma ED visit measure also developed in the PQMP and currently under review at NQF.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

None observed.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses 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 Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Appendix1_v1_Final_Measure_5.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University Hospitals Cleveland Medical Center
Co.2 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 617-699-3357-
Co.3 **Measure Developer if different from Measure Steward:** Collaboration for Advancing Pediatric Quality Measures (CAPQuaM)
Co.4 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 617-699-3357-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Role: Expert Panelists
Elizabeth Allen, M.D Ohio State University
Chitra Dinakar, M.D. Department of Pediatrics University of Missouri-Kansas City
Stephen Teach, M.D., M.P.H. Children's National Medical Center
Charles Macias M.D., M.P.H. Pediatrics, Baylor Texas Children's Hospital
Michael Cabana M.D., M.P.H. University of California, San Francisco

Barbara Yawn M.D., M.S. Olmstead Medical Center, University of Minnesota
Joan Connell M.D. University of North Dakota
Delaney Gracy M.D., M.P.H. The Children's Health Fund and Montefiore Children's Hospital
Sharlene Miner M.D. Emergency Medicine Inter Mountain Medical Center

ROLE: Steering Committee and Investigator
Wilson Pace, MD American Academy of Family Physicians – DARTNET Institute - University of Colorado
Lynn Olson, PhD American Academy of Pediatrics
Christina Bethell, PhD, MBA, MPH Child and Adolescent Health Measurement Initiative, Johns Hopkins University (Previous OHSU)
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Lawrence Kleinman, MD, MPH Icahn School of Medicine at Mount Sinai
Rebecca Anderson Mount Sinai Medical Center
Ira Nash, MD Northshore - Long Island Jewish Medical Center (previous Mount Sinai School of Medicine)
Eyal Shemesh, MD Icahn School of Medicine at Mount Sinai
Mary Barton, MD National Committee on Quality Assurance
Charles Homer, MD, MPH US Department of HHS (previous National Institute for Child Health Quality)
Marla Clayman, PhD American Institutes for Research (previous Northwestern University)
Foster Gesten, MD New York State Dept. of Health, Office of Health Insurance Programs
Jerod M. Loeb, PhD The Joint Commission

Robert Rehm National Committee for Quality Assurance
Steve Kairys, MD American Academy of Pediatrics/QuIIN
Erin DuPree, MD The Joint Commission (previous: Icahn School of Medicine at Mount Sinai)
David Baker The Joint Commission (Previous Northwestern)
Beverley Johnson, BSN* Institute for Patient- and Family-Centered Care
Doris Peter Consumers Union

ROLE: Senior Advisory Board Member and Investigator
Shoshanna Sofaer, DrPH American Institutes for Research (previous CUNY Baruch)
Lynne Richardson, MD Icahn School of Medicine at Mount Sinai
Ian Holzman, MD Icahn School of Medicine at Mount Sinai
Marilyn Kacica, MD New York State Dept. of Health, Division of Family Health

ROLE: Senior Advisory Board Member
Marc Lashley, MD Allied Pediatrics
Gary Mirkin, MD Allied Pediatrics
Barbara Kupferman, RN, CMCN, LNC AmeriChoice by United Healthcare
L. Gregory Pawlson, MD, MPH SE Healthcare QUALITY Consultants (previous Blue Cross Blue Shield)
John Santa, MD* (previous Consumers Union)
John Clarke, MD ECRI, PA Patient Safety Authority
Scott Breidbart, MD Empire Blue Cross Blue Shield/ Anthem
Jeff Terry, MBA GE Healthcare
Robert St. Peter, MD, MPH Kansas Health Institute
Ruth Stein, MD Montefiore Children's Hospital
Arthur Aufses, MD Icahn School of Medicine at Mount Sinai
Eric Rose, MD Icahn School of Medicine at Mount Sinai
Wendy Brennan, MS National Alliance on Mental Illness of New York City (NAMI - NYC Metro)
Laurel Pickering, MPH Northeast Business Group on Health
Tony Hope, BM BCh, PhD Oxford University, The Ethox Center
Martin Hatlie, JD Partnership 4 Pt. Safety, Consumers Advancing Patient Safety
Paul Wise, MD Stanford University
Mark Chassin, MD, MPH, MPP The Joint Commission
Ira Nash, MD Icahn School of Medicine at Mount Sinai
ROLE: Investigators/Key Staff
Rusty McLouth, MS AAFP
Kasey McCracken, MPH CAHMI
Elise Barrow, MPH Icahn School of Medicine at Mount Sinai
Kasey Coyne, BA Icahn School of Medicine at Mount Sinai
Natalia Egorova, PhD Icahn School of Medicine at Mount Sinai
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Melissa Manice, MPH Icahn School of Medicine at Mount Sinai
Joshua Oppenheimer, BA Icahn School of Medicine at Mount Sinai
Eve Oster, MD Pediatric Partners of Gwinnett (previous Icahn School of Medicine at Mount Sinai)
Daniel O'Sullivan, BS University of Virginia (previous Icahn School of Medicine at Mount Sinai)
Michael Parides, PhD Icahn School of Medicine at Mount Sinai
Barbara Rabin, MHA Icahn School of Medicine at Mount Sinai
Carolyn Rosen, MD Icahn School of Medicine at Mount Sinai
Melissa Saperstein, MSW Icahn School of Medicine at Mount Sinai
Eyal Shemesh, MD Icahn School of Medicine at Mount Sinai
Andrew Ting, MD Icahn School of Medicine at Mount Sinai
Adam Vella, MD Icahn School of Medicine at Mount Sinai
Virginia Walther, MSW Icahn School of Medicine at Mount Sinai

Jennifer Zeitlin, SD	Icahn School of Medicine at Mount Sinai, French Institute of Health and Medical Research (INSERM)
Marianne McPherson, PhD, MS	NICHQ
Joseph Anarella, MPH	NYS DOH, OHIP
Lee Sanders, MD, MPH	Stanford University
Kasey Coyne	Northwestern
Victoria Wagner	New York State Health Department Office Quality Patient Safety
Laura Soloway	New York State Health Department Office Quality Patient Safety
Kevin Malloy	New York State Health Department Office Quality Patient Safety
Judy Stribling	Icahn School of Medicine at Mount Sinai
Linda Radecki	AAP
Keri Thiessen	AAP – QuIIN
Laura Morris	New York State Department of Health
Olivia Lindly	OHSU
Kasey McCracken	OHSU
Sandeep Sharma	Icahn School of Medicine at Mount Sinai
Ann Nevar	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Suzanne Lo	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Amy Balbierz	Icahn School of Medicine at Mount Sinai
Samantha Raymond	Icahn School of Medicine at Mount Sinai
Allisyn Vachon	Icahn School of Medicine at Mount Sinai
Shannon Weber	Icahn School of Medicine at Mount Sinai
ROLE: Contractors	
Mary Fialkowski, MBA	The Joint Commission
Linda Hanold, MHSA	The Joint Commission
Elvira Ryan, MBA, BSN, RN	The Joint Commission
Sharon Sprenger, RHIA, CPHQ, MPA	The Joint Commission
Ann Watt, MBA, RHIA	The Joint Commission
Amy Slott	The Joint Commission
Denise Krusenoski	The Joint Commission
Elvira Ryan	The Joint Commission
Tasha Mearday	The Joint Commission

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:**

Ad.3 **Month and Year of most recent revision:**

Ad.4 **What is your frequency for review/update of this measure?**

Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2852
De.2. **Measure Title:** Optimal Asthma Control
Co.1.1. **Measure Steward:** MN Community Measurement
De.3. **Brief Description of Measure:** The percentage of pediatric (5-17 years of age) and adult (18-50 years of age) patients who had a diagnosis of asthma and whose asthma was optimally controlled during the measurement period as defined by achieving BOTH of the following:

- Asthma well-controlled as defined by the most recent asthma control tool result available during the measurement period
- Patient not at elevated risk of exacerbation as defined by less than two emergency department visits and/or hospitalizations due to asthma in the last 12 months

1d.3. **Developer Rationale:** Asthma control as demonstrated with the asthma patient reported outcome tool provides an indication of symptom control during a shorter term window, while the 12 month recall of emergency department visits and hospitalizations provides a longer term indication of control. The patients who achieve both targets are most effectively managing asthma symptoms.

S.4. **Numerator Statement:** The number of patients in the denominator whose asthma was optimally controlled during the measurement period as defined by achieving BOTH of the following:

- Asthma well-controlled as defined by the most recent asthma control tool result during the measurement period:
  -Asthma Control Test (ACT) greater than or equal to 20 (patients 12 years of age and older)
  -Childhood Asthma Control Test (C-ACT) greater than or equal to 20 (patients 11 years of age and younger)
  -Asthma Control Questionnaire (ACQ) less than or equal to 0.75 (patients 17 years of age and older)
  -Asthma Therapy Assessment Questionnaire (ATAQ) equal to 0 – Pediatric (5 to 17 years of age) or Adult (18 years of age and older).

AND

- Patient not at elevated risk of exacerbation as defined by less than two patient reported emergency department visits and/or hospitalizations due to asthma in the last 12 months

S.7. **Denominator Statement:** Patients aged 5 - 50 years at the start of the measurement period who were seen for asthma by an eligible provider in an eligible specialty face-to-face visit at least 2 times during the current or prior year measurement periods AND who were seen for any reason at least once during the measurement period.
S.10. **Denominator Exclusions:** Valid exclusions include patients who are nursing home residents, in hospice or palliative care, have died or who have COPD, emphysema, cystic fibrosis or acute respiratory failure.

De.1. **Measure Type:** Composite
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
S.26. **Level of Analysis:** Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

1d.1. **Composite Measure Construction:** all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)
Component Measures (if endorsed or submitted for endorsement):

New Measure -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

A version of this measure was previously reviewed as #1876, a 3-part composite, in the 2012-2013 Pulmonary Project. It was not recommended, but the previous Committee encouraged the developer to continue working on it. The measure is now a 2-part composite. As relevant, information related to #1876 is provided for context.

Summary of previous consideration:

#1876 had three components:

1. Asthma well-controlled as defined by the most recent asthma control tool result available during the measurement period
2. Patient is not at elevated risk of exacerbation as evidenced by patient-reported emergency department ED visits and in-patient hospitalizations due to asthma in the past 12 months. The total number of ED visits and hospitalizations due to asthma must be less than 2.
3. Patient has been educated about his or her asthma and self-management of the condition with a written asthma management plant present (created or reviewed and revised within the measurement period) that contains information about the patient's triggers, the patient's medication doses and effects of those medications, and what to do during an exacerbation.

The previous Committee commented as follows:

- Component 1. The surveys were developed and validated by performance testing in asthma clinics of allergists, which likely had higher degree of severity of asthma. Concern was expressed there could be significant numbers of patients with mild asthma in the broader population and whether the survey could adequately address these individuals.
- Component 2. The Committee questioned the evidence that ≤ 2 ED visits or hospitalizations means optimal control—e.g., with the measure's threshold, one hospitalization is considered well-controlled but two ED visits is not. Also of concern was that hospitalization and ED visits are of essentially equal weight, since they may be dramatically different events.
- Component 3. The Committee questioned why, as long as control is achieved (i.e., the other two components), why the action plan was identified as the method that must be included in order for a provider to be successful on the measure.
- Overall, the Committee was supportive of a composite and encouraged continued work, the measure failed on Evidence.

Summary of evidence:

The developer has dropped Component 3 from #1876; Component 1 and Component 2 remain essentially the same, with some edits in verbiage:

- This is an all-or-none composite that consists of two outcome measures (control and risk). The level of analysis is Clinican: Group/Practice.
- Based on the recommendations of three sets of clinical guidelines: the National Heart, Lung, and Blood Institute EPR-3 2007 (NHLBI), the Global Initiative for Asthma (GINA) updated in 2014, and again in April 2015, and the

Institute for Clinical Systems Improvement (ICSI) Asthma Guideline updated in 2012.
    - Asthma control and risk are both mentioned in the guidelines as important components for the management of patients with asthma. Both the GINA and NHLBI guidelines recommend that asthma control be the primary goal of asthma care and management [NHLBI evidence rating A = randomized control trials and rich body of data].
    - The rationale for the first component (asthma well-defined based on most recent assessment with tool) is provided: The use of validated control tools to monitor asthma control is mentioned by all three guidelines as appropriate means for assessing control. According to the GINA guidelines, "[The validated asthma control tools] have the potential to improve the assessment of asthma control, providing a reproducible objective measure that may be charted over time (week by week or month by month) and representing an improvement in communication between patient and health care professional." Information on a grade is not provided.
- Re: Component 2. The developer's rationale is that the 12-month recall of ED visits and hospitalizations provides a longer term indication of control. The developer does not report whether the new guidelines address a rationale for the second component. Neither is new research presented in this regard.
- In addition to guideline recommendations, the developers reviewed asthma medical research and found evidence that asthma control is correlated with improved health outcomes in patients with asthma.

Questions for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *Does the Committee wish to further discuss with the developer the rationale for the second component of the composite (ED/hospitalizations), given the lack of guidelines or other new empirical evidence since the last review?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following information:

ADULTS

Data collected in 2015, based on dates of service 7/1/2014 through 6/30/2015.
Number of clinics reportable (>= 30 patients): 436
Number of patients: 63,429
Mean = 49.4%
Standard deviation = 24.2%
Minimum = 0.0%
Maximum = 94.9%
IQR = 36.6%

10th	11.3%
20th	24.7%
30th	39.2%
40th	47.3%
50th	56.9%
60th	62.3%
70th	66.2%
80th	69.6%
90th	75.9%

CHILDREN

Data collected in 2015, based on dates of service 7/1/2014 through 6/30/2015.
Number of clinics reportable (>= 30 patients): 295
Number of patients: 39,408

Mean = 55.8%
Standard deviation = 22.5%
Minimum = 0.0%
Maximum = 95.5%
IQR = 28.0%
10th 22.2%
20th 35.5%
30th 48.5%
40th 56.1%
50th 60.7%
60th 66.1%
70th 71.0%
80th 74.2%
90th 79.6%

Disparities

The developer provides the following information:

ADULTS

- The 2014 statewide rate for Optimal Asthma Control - Adults is 47%
  Race: Native Hawaiian or Other Pacific Islander, 55%; White, 51%; Unknown racial group, 16%; Black or African American and American Indian or Alaska Native (% not given, only noted rate was significantly lower than statewide rate); Non-Hispanics, 49%; Hispanics, 42%
- English as preferred language, 47%; Somali as preferred language, 25%
- Patients born in South Korea, 61%; patients born in the United States, 47%. Patients born in Somalia had the lowest optimal rate at 30%; patients born in Mexico, 44%; 2 other country of origin optimal rates were significantly higher than the statewide average: India and the United States (rates not provided).

CHILDREN

- The 2014 statewide rate for Optimal Asthma Control - Children is 56%.
- Race: Asian racial group, 61%; White, 58%; American Indian or Alaskan Native racial group, 34%; African American, Other Race and Unknown groups had rates significantly below the statewide average (% not provided); Non-Hispanics, 56%; Hispanics, 53%.
- English as preferred language, 56%; Somali as preferred language, 52%
- Patients born in United States, 56%; patients born in Somalia had the lowest optimal rate, 42%; patients born in Mexico, 44%.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

1c. Composite - Quality Construct and Rationale
Maintenance measures – same emphasis on quality construct and rationale as for new measures.

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

The developer provides the following:

- This composite measure is a patient level all-or-none composite.
- The composite rationale is supported by the asthma patient reported outcome tool, which provides an indication of symptom control during a shorter term period, while the 12-month recall of ED visits and hospitalizations is indicative of longer-term control.
- The developers calculate the numerator at the patient level and define numerator compliance as the patient achieving both components of the measure. The numerator is calculated with equal weighting between the two

components.

Questions for the Committee:

- *Are the quality construct and a rationale for the composite explicitly stated and logical?*
- *Is the method for aggregation and weighting of the components explicitly stated and logical?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

***1a. Evidence** to Support Measure Focus*

Comments:

**2-part composite measure. Previous failure on evidence with last review. Lack of guidelines or other new evidence supplied by the developer.

**This measure draws heavily on clinical guidelines, but the measure itself has not been directly subject to evidence-based review.

**This is a composite measure is looking ate overall level of asthma control and level of risk. The composite measure was revised from 3 components to a total of 2 components. The third component of having an asthma action plan in place was dropped as if the patient reported good control, the need for a specific plan was not needed to be considered acceptable. This is an all-or-nothing measure. This measure is directly related to the desired outcome of improved asthma control. There is still not great evidence for Component 2 (12-month recall of ED visits and hospitalizations as a longer term indication of control). We should talk with the developer about why component 2 was included despite lack of new empirical evidence.

**Composite measure that is composed of two component measures; some graded clinical evidence for component one; no graded evidence for component two for guidelines, but some evidence found for asthma control and correlation with improved health outcomes for patients with asthma.

1b. Performance Gap

Comments:

**Performance data provided with a mean of 49.4% in adults. Data on subgroups provided and demonstrates variation in mean.
**Performance Gap is dramatic for both adults and children based on data gathered in Minnesota among a large number of patients and clinics. The data provided also demonstrates disparity in care delivery.
**There does appear to be a gap that warrants a national performance measure. In adults, statewide rate of optimal asthma control is 47% while in children the optimal asthma control rate is 56%. There do appear to be ethic/racial gaps as well.
**Developer analyses indicates a gap in performance for the composite; some disparities noted based on race, language, ethnic origin.

1c. High Priority (previously referred to as High Impact)

Comments:

**The rationale for component 2 should be discussed in terms of evidence. The developer lays out the composite quality construct and rationale.
**The composite methodology is simply an "all or nothing" measure for the 2 following asthma measures: 1) "Asthma well controlled as defined by the most recent asthma control tool available during the measurement period" and 2) Patient had less than 2 ED or hospital visits in in the past 12 months. Unlike other composite measures (including other measures developed by MN Community Measurement), neither of the sub-components is NQF endorsed. The composite is not robust given the weakness of the underlying measures that compose it.
**This composite performance measure is using a short term measure (patient reported outcome tool) and a long term measure (12-month recall of ED visits and hospitalizations). There is equal weighting between the measures and the rationale is logical. They are both self-report measures. The quality construct is not explicitly stated.
**The developer addressed all of the aforementioned sections of the quality construct in a logical and complete manner.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Electronic Clinical Data, Electronic Clinical Data: Electronic Health Record, Paper Medical Records

Specifications:

- The numerator of this measure is the number of patients in the denominator whose asthma was optimally controlled during the measurement period as defined by achieving BOTH of the following:
  - Asthma well-controlled as defined by the most recent asthma control tool result during the measurement period:
    -Asthma Control Test (ACT) greater than or equal to 20 (patients 12 years of age and older)
    -Childhood Asthma Control Test (C-ACT) greater than or equal to 20 (patients 11 years of age and younger)
    -Asthma Control Questionnaire (ACQ) less than or equal to 0.75 (patients 17 years of age and older)
    -Asthma Therapy Assessment Questionnaire (ATAQ) equal to 0 – Pediatric (5 to 17 years of age) or Adult (18 years of age and older).

    AND
  - Patient not at elevated risk of exacerbation as defined by less than two patient reported emergency department visits and/or hospitalizations due to asthma in the last 12 months
- The denominator of this measure is: *Patients aged 5-50 years at the start of the measurement period who were seen for asthma by an eligible provider in an eligible specialty face-to-face visit at least 2 times during the current or prior year measurement periods AND who were seen for any reason at least once during the measurement period*.
- The ICD-9 and ICD-10 codes have been included in the specification details.
- The calculation algorithm is stated in S.18.
- The measure is risk model has 1 variable: (Commercial, Medicare, Medicaid, Self-Pay)

Questions for the Committee :

- *Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers

- The developers used testing data from 7/1/2013 – 6/30/2014
- Data sample:
  - ADULTS: 415 clinics (All); 59,717 patients (All)
  - CHILDREN: 291 clinics (All); 36,666 patients (All)

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Reliability was assessed using the beta-binomial approach (BETABIN/ SAS).

Results of reliability testing

- ADULTS: Reliability = 0.972
- CHILDREN: Reliability = 0.951

- Distribution of scores is provided for both populations.
- The developer states the measure construct is reliable, citing a reliability score of 0.7 or greater as the threshold.

Guidance from the Reliability Algorithm: 1→ 2→ 4 → 5 → 6 (eligible for HIGH rating)

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence? Specifically, are the specifications for the second component consistent with the evidence (patient not at elevated risk of exacerbation as defined by less than two emergency department visits and/or hospitalizations due to asthma in the last 12 months)?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Composite performance measure score empirical validity testing:
- The developer tested the correlation of medical groups' performance with their performance on the Optimal Diabetes Care measure (NQF #0729), hypothesizing that the quality of care provided by a medical group to patients with diabetes would be of similar quality as the care provided to patients with asthma, and the respective performance measure scores should demonstrate such.
- Validity testing for critical data elements
- Completed in four steps: denominator certification, data quality checks, validation audit, and the two-week medical group review period. Data quality checks in place for evaluating the accuracy of data submitted.

Validity testing results:

The developer reports the following

- Composite performance measure score empirical validity testing (based on linear regression analysis):
  - The correlation coefficient was 0.63 for the adult measure, which the developer states represents a "fairly strong correlation" between the medical group's performance on the Optimal Diabetes Care measure and the Optimal Asthma Control (Adults) measure.
  - The correlation coefficient was 0.66 for the children measure, which the developer states represents a "fairly strong correlation" between the medical group's performance on the Optimal Diabetes Care measure and its performance on the Optimal Asthma Control (Children) measure.
- Validity testing for critical data elements:
- The developer states 100% of groups achieved the desired > 90% data accuracy when submitted data was

compared to medical record data (EMR or paper) of the patient. The developer does not provide specific data on each critical data element, as requested by NQF guidance, but rather presents results on a record basis as noted below.

Individual critical data element validation results (% agreement with true source)

Critical data elements	# of records with errors that do NOT alter result	% of records with errors that do NOT alter result	# of records with errors that alter result	% of records with errors that alter result	Total # audited
DOB	2	0.16%	0	0.00%	
Control test date	9	0.74%	1	0.08%	
Control test result	7	0.57%	0	0.00%	
ED/Hosp date	11	0.90%	7	0.57%	
ED #	5	0.41%	7	0.57%	
Hosp #	5	0.41%	7	0.57%	
Unique records	18	1.47%	8	0.66%	1221
Unique entities	5	4.27%	3	2.56%	117

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following:

- The following exclusions must be applied to the eligible population:
  - Patient had a diagnosis of cystic fibrosis, COPD, emphysema or acute respiratory failure (*Obstructive Lung and Respiratory Failure* Value Set)
- The following exclusions are allowed to be applied to the eligible population:
  - Patient was a permanent nursing home resident at any time during the measurement period
  - Patient was in hospice or receiving palliative care at any time during the measurement period
  - Patient died prior to the end of the measurement period
  - Documentation that diagnosis was coded in error
- Per the developer, the impact of the exclusions is small, but the majority of exclusions are optional exclusions. The developer states the volume of excluded patients is consistently less than 1% of the total population in both the adult and children populations.

Questions for the Committee:

- *Are "optional" exclusions problematic for a national performance measure for accountability purposes?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

- The developer appears to have analyzed standard demographic variables it had available—gender, age, zip, race/ethnicity, country of origin, primary language and insurance product—as well as three clinical variables identified by an expert panel (comorbidity of depression, tobacco use and tobacco exposure). We note the measure information form, testing form, and attachments slightly differ and should be reconciled, although the conclusion as noted next is clear.
- The developer reports the effects of all of the candidate risk adjusters except insurance product were not statistically significant. Patients who are MSHO, Medicaid, Special Needs, Uninsured or self-pay are 37% less likely than patients with commercial insurance to have optimal asthma care; it confirmed separately to NQF staff that the final measure includes only insurance product.
- The developer presented the following results:
    - At the patient level, the average optimal asthma control (OAC) was 26.2% (standard deviation = 23.6). The average number of patients reported by a clinic was 139 (standard deviation = 121). The average age in the examined population was 35.2 years. Within the population, 66% were female, 68.9% had commercial insurance, 5.72% had Medicare coverage, and 15.5% had Medicaid coverage.
    - There was significant heterogeneity across clinics in insurance product mix ($\chi2 = 13{,}309$, $p < .001$), patient age ($\chi2 = 2{,}798$, $p < .001$), gender ($\chi2 = 1035$, $p < .001$), tobacco use ($\chi2 = 82539$, $p < .001$), second hand tobacco exposure ($\chi2 = 95100$, $p < .001$) and distance to the clinic ($\chi2 = 27{,}061$, $p < .001$).
- The developer tested the overall correlation between the unadjusted and risk adjusted OAC measure using the Pearson correlation and a Kendall's Tau correlation. The Pearson correlation equaled 0.99. The Kendall's Tau correlation was 0.93. The developer concludes both approaches show a "very strong" correlation between the unadjusted and adjusted measure.
- The developer also used the various methods to compare the effects of risk adjustment on clinic rank (table 2 and table 3). These analyses showed no major difference between clinic rankings by decile or risk adjustment.
- In the section 2b.5 on Meaningful Differences, the developer also reports the following: "*The risk adjustment is done using an Actual to Expected methodology. This methodology does not alter the result of a clinic/medical group; the actual rate remains unchanged but instead of comparing the rate to the raw market average, a unique expected rate based on the proportion of each risk category for each clinic/medical group is used for comparison.*

Questions for the Committee:

- *Is the risk adjustment strategy—statistical model with 1 variable—appropriate for a national performance measure?*
- *Is the test sample of data from a single state adequate to generalize a risk model for widespread implementation?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

For Meaningful Differences, the developer reports the following:

- "The risk adjustment is done using an Actual to Expected methodology. This methodology does not alter the result of a clinic/medical group; the actual rate remains unchanged but instead of comparing the rate to the raw market average, a unique expected rate based on the proportion of each risk category for each clinic/medical group is used for comparison."
    - The comparison between clinics is the aforementioned test of significance and the actual to expected ratio: actual percentage of patients meeting criteria divided by the expected percentage of patients meeting criteria for the particular entities mix of patient risk."
    - The developer used a one population proportions test to determine whether there was a statistically significant difference between the expected rate and the actual rate achieved by a clinic/medical group. The methodology used a 99% test of significance.
    - In a separate communication, the developer stated the actual to expected comparison approach also is used for reporting . The performance score in not adjusted, but the actual to expected (based on composition of patients attributed to that facility) rates are reported. The developer states more than 50% of groups pass the test for significance to be meaningfully different than the expected rate for their patient population.

Question for the Committee:

- *Does this methodology demonstrate meaningful differences **across** measured entities?*
- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

The developer notes the following:

- For this measure, elements missing from any component are counted as a numerator component fail and therefore the patient would be accounted for and remain in the denominator.
- The developer states missing data are not a factor for this measure, but does not quantify its occurrence

2d. Composite measure: construction

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

- The developer considered including the following components: an asthma management plan and avoidance of tobacco exposure. Both components were removed from consideration.
  - Asthma management was removed due to inconsistent evidentiary support
  - Exposure to tobacco was removed due to the physician's limited ability to influence the behaviors of individuals in the home
- No empirical analysis was performed on the composite construction. The developer provides the following rationale for construction: included the patient outcome component to demonstrate short term control and included the number of ED visits and hospitalization to demonstrate long term control.
- No empirical analysis was provided demonstrating that the aggregations and weighting rules are consistent with the construct and achieve the objective of simplicity to the extent possible, nor was a justification provided for the lack of empirical analysis, as requested by NQF.
- The developer did not identify the aggregation and weighting rules that were considered and the pros and cons of each, as requested by NQF when statistical results are not available from empirical analysis.
- No rationale was provided for the equal weighting of the two components.

Questions for the Committee:

- *Do the component measures fit the quality construct?*
- *Are the objectives of parsimony and simplicity achieved while supporting the quality construct?*

Guidance from the Validity Algorithm: 1→ 2→ (Insufficient) OR → 3 → 6 → 7 → 8 (eligible for HIGH rating)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**We should discuss whether the specifications for the second component consistent with the evidence.

**It is compelling that 100% of groups achieved the 90% data accuracy when submitted data was compared with the EMR or paper chart.

**Developer notes that specifications are somewhat inconsistent with the evidence but does not offer specific detail on shortcomings.

2a2. Reliability Testing

Comments:

**Results of validity testing demonstrate sufficient validity so that conclusions about quality.

**Validity was based on a correlation between clinics performance on the Optimal Diabetes Care measure to the Optimal Asthma Control measure(s). The assumption being the high performance on one measure should be correlated with the other. This assumption appears tenuous plus correlation coefficient of 0.62 (adults) and 0.66(children) are weak.

**Validity testing was completed and the analysis looks to be a fairly strong correlation between the Optimal Asthma Care measure

as with other measures including the Optimal Diabetes control.
**Developer notes that total number of records audited were 1221, with 117 unique entities. Validity was tested at the measure score levels with empirical validity testing; correlation was tested against performance on Optimal Diabetes Care measure that would be similar to asthma measure.

2b2. Validity Testing
Comments:
**For this measure, elements missing from any component are counted as a numerator component fail and therefore the patient would be accounted for and remain in the denominator. The developer states missing data are not a factor for this measure, but does not quantify its occurrence
**Exclusions are appropriate. Risk adjustment includes demographic variables and comorbidities of depression, tobacco use, and tobacco exposure yet with the exception of insurance product, none demonstrated statistical significance. The risk adjustment model does not appear to be robust to explaining variation in outcomes. It is not clear why the developer includes a risk adjustment model. Missing data does not appear to be an issue.
**The developer states that missing data are not a factor for this measure, but does not quantify its occurrence. I think this needs to be more clearly explained.
**Developer notes that missing data would not be a factor for this measure.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**Results appear to demonstrate sufficient reliability to determine differences in performance.
**Reliability appears to be adequate to reliability differentiate performance among providers.
**I don't think that the inclusion pose a serious threat to the reliability of the national performance measure. I do have some concerns that we are excluding patients with COPD, emphysema, and acute respiratory failure as patients live with comorbidities and it would be more representative to collect the data on all of the patients and then stratify for risk.
**Total data sample for children and adults noted at 706 clinics and greater than 96K patients total. Testing was conducted at measure score with data source level of analysis. Minimum reliability score set at 0.7; for adults and children, reliability levels significantly above level.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic health records (EHRs)
- The developer indicates there are no fees, licensing, or other requirements to use any aspect of the measure, unless practices wish to use the ACQ, ATAQ or ACT beyond the current agreement.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments:
**All data elements are in defined fields in electronic health records (EHRs). Data elements are likely routinely generated during care deliver.
**All data elements are routinely generated in EHRs during the course of care delivery.
**Are these measures currently being used nationally? If not, how hard are they to implement? Are they currently the standard of

practice?
**Developer notes that all elements of the measure are readily available through EHRs

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
Public Reporting
- MNHealthScores: http://www.mnhealthscores.org/asthma-children-842
- MNHealthScores:http://www.mnhealthscores.org/medical-group-measure-detail/adults/#/results

Payment Program
- PQRS

Regulatory and Accreditation Programs
- MN Health Care Homes

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
- MN Health Care Homes

Improvement results
- The developer notes the measure was revised in Report Year 2015 and does not have comparable historical data to demonstrate progress on improvement.

Potential harms
- The developer notes no unintended consequences.

Feedback:
- No feedback provided on QPS. Measure reviewed by MAP in 2013-2014. MAP 2014 decisions are as follows:
  - Medicare Shared Savings Program: Support;
  - Physician Quality Reporting System (PQRS): Support;
  - Physician Compare: Conditional Support;
  - Value-Based Payment Modifier Program: Conditional Support
- MAP provided the following feedback:
  - *Composite outcome measure important to consumers. The measure has been revised to address concerns raised by NQF Steering Committee. Conditional on revised measure being submitted to NQF. Relative improvement for severe asthmatics is not included. Minnesota uses comparisons of like providers, i.e., pulmonologist compared to each other, etc. Upper age limit of 50 years may not be warranted as increasing number of older patient have asthma.*

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Reported online currently, publicly available.
**The data is publicly reported in Minnesota. The measure has great potential to improve care, though evidence is limited. The risk of unintended consequences are minimal.
**Currently used in Minnesota for public reporting, payment programs, regulatory and accreditation programs, and quality improvement with benchmarking. The developers do not cite any unintended consequences of using this measure. I am curious about the burden for providers when collecting this data and the burden of abstracting the data. Otherwise, I don't see any unintended consequences.
**The measure is currently being publicly reported and used in accountability programs: MNHealthScores; PQRS; MN Health Care Homes

Criterion 5: Related and Competing Measures

Related or competing measures

- 2794: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma (University Hospitals Cleveland Medical Center)
- 2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Click here to enter measure title

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: Click here to enter a date

Instructions

- *For composite performance measures:*

- *A separate evidence form is required for each component measure unless several components were studied together.*
- *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*

- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Click here to name the health outcome

☒Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

The Optimal Asthma Control measure is an all or none composite that consists of two outcome measures (control and risk). They draw heavily on the recommendations of three sets of clinical guidelines: the National Heart, Lung, and Blood Institute EPR-3 2007 (NHLBI), the Global Initiative for Asthma (GINA) updated in 2014, and again in April 2015, and the Institute for Clinical Systems Improvement (ICSI) Asthma Guideline updated in 2012. Asthma control is stated by clinical guidelines to be the primary goal of asthma therapy. The use of asthma control identifies a patient´s level of impairment due their condition. The assessment of risk is included in the guidelines as an important complimenting assessment of overall asthma control based on past experience. It is considered a separate determining point from the level of impairment due to asthma.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Asthma control and risk are both mentioned in the guidelines as important components for the management of patients with asthma. For asthma control, both the GINA and NHLBI guidelines recommend that asthma control be the primary goal of asthma care and management [NHLBI evidence rating A = randomized control trials and rich body of data]. The use of validated control tools to monitor asthma control is mentioned by all three guidelines as appropriate means for assessing control. According to the GINA guidelines, "[The validated asthma control tools] have the potential to improve the assessment of asthma control, providing a reproducible objective measure that may be charted over time (week by week or month by month) and representing an improvement in communication between patient and health care professional." The NHLBI guidelines suggest that asthma control should be the impetus for adjusting asthma treatment. In addition to guideline recommendations, a review of asthma medical research during the development of the Optimal Asthma Control measure found evidence that asthma control is correlated with improved health outcomes in patients with asthma.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report
Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form Optimal_Asthma_Control_evidence_attachment.docx
1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

ADULTS
Analysis on data collected in 2015, based on dates of service 7/1/2014 through 6/30/2015.
Number of clinics reportable (>= 30 patients): 436
Number of patients: 63,429
Mean = 49.4%
Standard deviation = 24.2%
Minimum = 0.0%
Maximum = 94.9%
IQR = 36.6%

10th	11.3%
20th	24.7%
30th	39.2%
40th	47.3%
50th	56.9%
60th	62.3%
70th	66.2%
80th	69.6%
90th	75.9%

CHILDREN
Analysis on data collected in 2015, based on dates of service 7/1/2014 through 6/30/2015.
Number of clinics reportable (>= 30 patients): 295
Number of patients: 39,408
Mean = 55.8%
Standard deviation = 22.5%
Minimum = 0.0%
Maximum = 95.5%
IQR = 28.0%

10th	22.2%
20th	35.5%
30th	48.5%
40th	56.1%
50th	60.7%
60th	66.1%
70th	71.0%
80th	74.2%
90th	79.6%

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities*

include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.

ADULTS

The 2014 statewide rate for Optimal Asthma Control - Adults is 47%. By race,the Native Hawaiian or Other Pacific Islander racial group had the highest rate of optimal control at 55%; however, this rate was not significantly above the statewide average (47%) or the White racial group optimal rate (51%). The White racial group was the only racial group significantly above the statewide average at 51%. The Unknown racial group had the lowest rate of optimal control at 16%, and this was significantly below the statewide average. Additionally
the Some Other Race, American Indian or Alaskan Native and the Black or African American groups had rates significantly below the statewide average as well.
Notably, the Black or African American racial group and American Indian or Alaska Native group had significantly lower optimal rates than the Asian and
White racial groups.

Non-Hispanics had a significantly higher rate of optimal control (49%) than Hispanics (42%) and the statewide average (47%). Hispanics had rate of optimal control that was significantly lower than the statewide average.

Patients who indicated English was their preferred language had the highest rate of optimal control (47%), and this rate was significantly higher than the statewide average (47%). The lowest rate of optimal control was held by patients that indicated Somali as their preferred language at 25%; this rate was significantly below the statewide average. The English preferred language optimal rate was significantly higher than both the Somali and Spanish referred language optimal rates.

Patients born in South Korea had the highest rate of optimal control at 61%, and this rate was significantly higher than the statewide average (56%) and the optimal rate for patients born in the United States (47%). Additionally, two other country of origin optimal rates were significantly higher than the statewide average: India and the United States. Patients born in Somalia had the lowest optimal rate at 30%, which was significantly lower than the statewide average. The optimal rate for patients born in Mexico (44%) was also significantly lower than the statewide average. Notably, patients born in Mexico and Somalia had rates of optimal control below the statewide average and the United States-born optimal rate.

CHILDREN

The 2014 statewide rate for Optimal Asthma Control - Children is 56%.

By race, the Asian racial group had the highest rate of optimal control at 61%, which is significantly above the statewide average (56%). However, it was not significantly above the White racial group, which was the only other racial group significantly above the statewide average at 58%. The American Indian or Alaskan Native racial group had the lowest rate of optimal control at 34%, and this was significantly below the statewide average. Additionally, the Black or African American, Some Other Race and the Unknown groups had rates significantly below the statewide average as well. Notably, the Black or African American racial group had a significantly lower optimal rate than the Asian and White racial groups.

Non-Hispanics had a significantly higher rate of optimal control (56%) than Hispanics (53%), but this rate was not significantly higher than the statewide average (56%). Hispanics had rate of optimal control that was significantly lower than the statewide average.

Patients who indicated English was their preferred language had the highest rate of optimal control (56%), but this rate was not significantly higher than the statewide average (56%). There were no optimal control rates for preferred language that were significantly higher than the statewide average. The lowest rate of optimal care was held by patients that indicated Somali as their preferred language at 52%; this rate was significantly below the statewide average. The English preferred language optimal rate was significantly higher than the Somali preferred language optimal rate.

Patients born in United States had the highest rate of optimal control at 56%; however, this rate was not significantly higher than the statewide average (56%). Patients born in Somalia had the lowest optimal rate at 42%, which was significantly lower than the statewide average as well as the optimal rate for patients born in Mexico (44%). Notably, patients born in Mexico and Somalia had rates of optimal control below the statewide average and the United States-born
optimal rate.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Asthma is one of the most common chronic diseases in the United States. An estimated 18.7 million adults and 7.0 million children have asthma. In 2008, asthma accounted for 456,000 hospitalizations and 1.8 million emergency department visits. Adults and children with asthma experienced 14.2 million missed work days and 10.5 million missed school days, respectively, in 2008. Asthma was associated with $56 billion in total costs in the U.S. in 2007.

Asthma affects an estimated 392,000 Minnesota children and adults who currently have the disease. In Minnesota, 90,000 children (7.0%) currently have asthma, and over 300,000 adults (7.6%).
In 2010, there were nearly 20,000 emergency department visits in Minnesota due to asthma and more than 3,500 hospitalizations.
In Minnesota, it is estimated that, in 2004, asthma cost $240 million directly in hospitalizations, emergency department visits, office visits, and medications, and $181 million indirectly in lost school and work days, for total of $421 million.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Akinbami LJ, Moorman JE, Bailey C, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. NCHS data brief, 2012, No 94. Hyattsville, MD: National Center for Health Statistics.

Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005–2009. National Health Statistics Reports; 2011, No 32. Hyattsville, MD: National Center for Health Statistics.

Barnett SB and TA Nurmagambetov. Costs of asthma in the United States: 2002-2007. J Allergy Clin Immunol 2011;127:145-52.

Asthma in Minnesota: 2012 Epidemiology Report. Minnesota Department of Health. St. Paul, MN. June 2012.

Asthma care quality improvement: A resource guide for state action, Agency for Healthcare Research & Quality (AHRQ), 2006. http://www.ahrq.gov/qual/asthmacare/asthmatab1_3.htm

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

From the NHLBI Guidelines for the Diagnosis and Management of Asthma
An assessment of the impairment domain for determining the severity of disease (in patients on no long-term-control treatment before treatment is initiated) or the level of control (after treatment is selected) usually can be elicited by careful, directed history and lung function measurement. Standardized questionnaires like the Asthma Control Test (ACT) (Nathan et al. 2004), the Childhood Asthma Control Test (Liu et al. 2007), the Asthma Control Questionnaire (Juniper et al. 1999b), the Asthma Therapy Assessment Questionnaire (ATAQ) control index (Vollmer et al. 1999), and others have been developed to facilitate and standardize the assessment of the impairment domain of asthma control.

Validity of the ACQ was confirmed by comparison to Asthma Quality of Life Questionnaire (AQLQ) and the Medical Outcomes Survey Short Form-36 (SF-36), clinician global rating of change (based on consultation with the patient, spirometry, 1-week diary of short-acting ß2-agonist use and morning prebronchodilator peak expiratory flow, AQLQ and SF-36 data).

The ACT includes an overall self-assessment of asthma control as well as an assessment of the frequency of shortness of breath and general asthma symptoms, use of rescue medications, and the effect of asthma on daily functioning.

The ATAQ also include a self assessment of overall control as well as an indication of missed work, school or normal daily activity, wakefulness at night and use of recue medications.

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
  - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
  - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

This composite measure is a patient level all-or-none composite in which the desired goal is for the patient to achieve control of their asthma symptoms as demonstrated by:
1. asthma well-controlled as defined by the most recent asthma control patient reported outcome tool result available during the measurement period
2. patient not at elevated risk of exacerbation as defined by less than two emergency department visits and/or hospitalizations due to asthma in the last 12 months.

The numerator is calculated at the patient level and numerator compliance is defined as the patient achieving both components of the measure. The components are treated equally.

1d.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**
Asthma control as demonstrated with the asthma patient reported outcome tool provides an indication of symptom control during a shorter term window, while the 12 month recall of emergency department visits and hospitalizations provides a longer term indication of control. The patients who achieve both targets are most effectively managing asthma symptoms.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**
The numerator is calculated at the patient level with equal weighting between the two components.

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care : Asthma

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Functional Status, Health and Functional Status, Health and Functional Status : Functional Status, Patient and Family Engagement

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://mncm.org/wp-content/uploads/2013/04/Optimal-Asthma-Control-2015_Asthma-Education-and-Self-Management-2015-Data-Collection-Guide-FINAL-v1.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** MNCM_Data_Dictionary_Optimal_Asthma_Control.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of patients in the denominator whose asthma was optimally controlled during the measurement period as defined by achieving BOTH of the following:

- Asthma well-controlled as defined by the most recent asthma control tool result during the measurement period:
  -Asthma Control Test (ACT) greater than or equal to 20 (patients 12 years of age and older)
  -Childhood Asthma Control Test (C-ACT) greater than or equal to 20 (patients 11 years of age and younger)
  -Asthma Control Questionnaire (ACQ) less than or equal to 0.75 (patients 17 years of age and older)
  -Asthma Therapy Assessment Questionnaire (ATAQ) equal to 0 – Pediatric (5 to 17 years of age) or Adult (18 years of age and older).

AND

- Patient not at elevated risk of exacerbation as defined by less than two patient reported emergency department visits and/or hospitalizations due to asthma in the last 12 months

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
1 year

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Asthma control test date
Enter the date of the most recent asthma control test on or prior to 06/30/2015.
Leave BLANK if an asthma control test was never performed.

- Do NOT enter any test date that occurred after 06/30/2015. A date after the measurement period will create an ERROR upon submission.
- Enter the date of the visit, telephone call, e-visit or other contact during which the asthma control test was administered (e.g., a test administered to the patient via phone).
- Test from another provider is acceptable (not required) if documented in the reporting clinic's record and is more recent than the reporting clinic's test.
- The following are approved, valid asthma control tests and must be giving according to validated age ranges. Age should be calculated as the date the asthma control test was administered. Tests other than the ones listed below will not be accepted.
  - ACT (Asthma Control Test); valid for patients 12 and older.

- CACT (Child-Asthma Control Test); valid for patients 11 and younger.
- ACQ (Asthma Control Questionnaire); valid for patients 17 and older.
- ATAQ (Asthma Therapy and Assessment Questionnaire); valid for patients 5 to 50.

Asthma control test name
Enter a code to indicate the most recent asthma control test (on or prior to 06/30/2015) given to the patient using the codes below. This test name should correspond to the test given on the date in Column U.
Leave BLANK if an asthma control test was never performed.
Leave BLANK if the wrong test was administered to the patient at the visit (e.g., a 12-year-old patient received the C-ACT instead of the ACT).
1 = Asthma Control Test (ACT)
2 = Child-Asthma Control Test (C-ACT)
3 = Asthma Control Questionnaire (ACQ)
4 = Asthma Therapy Assessment Questionnaire (ATAQ)
- The test used will be validated using the patient's date of birth and the date the test was given.

Asthma control test score
Enter the score of the most recent asthma control test (on or prior to 06/30/2015). The score should correspond to the test date listed in Column U and to the test name listed in Column V.
Leave BLANK if no control tests exist.
Leave BLANK if the wrong test was administered to the patient (e.g., a 12-year-old patient received the C-ACT instead of the ACT).
- If the test score is blank or not complete, look for an earlier completed asthma control test completed within the measurement period. Update Column U and Column V to reflect the new test date and name.
- Do NOT submit partial or incomplete scores. If there is not a test in the record with a complete score, leave Columns U, V and W blank.

Date of patient reported hospitalizations and emergency department visits
Enter the most recent date within the measurement period that the patient is asked about any hospitalizations and emergency department visits.
Leave BLANK if the patient was not asked about hospitalizations and emergency department visits. A date is necessary for rate calculation. Do NOT leave blank unless there is no data.
- This date must be associated with the patient-reported emergency department and hospitalizations columns during the past 12 months (Columns Y and Z).

Do NOT enter any visit that occurred after 06/30/2015. A date after the measurement period will create an ERROR upon submission.

Number of emergency department visits due to asthma that did NOT result in a hospitalization in the past 12 months (from date of visit)
Enter a numeric value for the number of emergency department (ED) visits due to asthma as stated by the patient (e.g. 0, 1, 2, etc.). Do NOT include urgent care visits.
Leave BLANK if the patient was not asked about emergency department visits or there is no data.
0 = Patient reports "0" or had no ED visits
1= Patient reports "1" ED visits
2= Patient reports "2" ED visits; etc.
A value is necessary for rate calculation. Do NOT leave blank unless there is no data. Enter the value collected and recorded asked and documented on or prior to 06/30/2015. Do NOT enter a number recorded prior to 07/01/2014.
- The patient should respond with a number of visits for the prior 12 months regardless of when the visit occurs – if the visit occurs in September of 2014, the previous 12 months would be September 2013 to August 2014. If the visit occurs in January 2015, the previous 12 months would be January 2014 to December 2014.
- Do NOT search for actual emergency department visits in your record system. This value must reflect what the patient reported when asked.
- If using an EMR, consider building a field to capture this data. If using paper, check the progress notes and other documentation from the most recent visit looking backwards.
- To be included in the numerator, the total number of BOTH emergency department visits AND inpatient hospitalizations due to asthma must equal ZERO or ONE.

Number of inpatient hospitalizations due to asthma during the past 12 months (from date of visit)
Enter a numeric value for the number of emergency department visits due to asthma as stated by the patient (e.g. 0, 1, 2, etc.).

Leave BLANK if patient was not asked about hospitalizations or there is no data
0 = Patient reports "0" or had no hospitalizations
1= Patient reports "1" hospitalization
2= Patient reports "2" hospitalizations; etc.
A value is necessary for rate calculation. Do NOT leave blank unless there is no data. Enter the value collected and recorded and documented on or prior to 06/30/2015. Do NOT enter a number recorded prior to 07/01/2014.

- Enter the patient reported number of inpatient hospitalizations due to asthma. The patient should respond with a number of visits for the prior 12 months regardless of when the visit occurs – if the visit occurs in September of 2014, the previous 12 months would be September 2013 to August 2014. If the visit occurs in January 2015, the previous 12 months would be January 2014 to December 2014.
- Do NOT search for actual hospitalizations in your record system. This value must reflect what the patient reported when asked.
- If using an EMR, consider building a field to capture this data. If using paper, check the progress notes and other documentation from the most recent visit looking backwards.
- To be included in the numerator, the total number of BOTH emergency department visits AND inpatient hospitalizations due to asthma must equal ZERO or ONE.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients aged 5 - 50 years at the start of the measurement period who were seen for asthma by an eligible provider in an eligible specialty face-to-face visit at least 2 times during the current or prior year measurement periods AND who were seen for any reason at least once during the measurement period.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Children's Health

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patients who meet each of the following criteria are included in the population:

- Patient was age 5 to 50 years at the start of the measurement period (date of birth was on or between 07/01/1964 to 07/01/2009).
  - Age 5 to 17 years at the start of the measurement period (date of birth was on or between 07/01/1997 to 07/01/2009).
  - Age 18 to 50 years at the start of the measurement period (date of birth was one or between 07/01/1964 to 06/30/1997).
- Patient was seen by an eligible provider in an eligible specialty face-to-face visit at least two times during the last two measurement periods (07/01/2013 to 06/30/2015) with visits coded with an asthma ICD-9 code (in any position, not only primary). Use this date of service range when querying the practice management or EMR system to allow a count of the visits.
- Patient was seen by an eligible provider in an eligible specialty face-to-face visit at least one time during the measurement period (07/01/2014 to 06/30/2015) for any reason. This may or may not include a face-to-face visit with an asthma ICD-9 code.
- Diagnosis of asthma; ICD-9 diagnosis codes include: 493.00 to 493.12, 493.81 to 493.92.

Eligible specialties: Family Practice, General Practice, Internal Medicine, Pediatrics, Allergy/Immunology, and Pulmonology.

Eligible providers: Medical Doctor (MD), Doctor of Osteopathy (DO), Physician Assistant (PA), Advanced Practice Registered Nurses (APRN).

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Valid exclusions include patients who are nursing home residents, in hospice or palliative care, have died or who have COPD, emphysema, cystic fibrosis or acute respiratory failure.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patient was a permanent nursing home resident during the measurement period.
Patient was in hospice or palliative care at any time during the measurement period.
Patient died prior to the end of the measurement period.
Documentation that diagnosis was coded in error.
Patient has COPD (codes 491.2, 493.2x, 496, 506.4)

Patient has emphysema (codes 492, 506.4, 518.1, 518.2)
Patient has cystic fibrosis (code 277.0)
Patient has acute respiratory failure (code 518.81)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Patient age group (children 5-17 years, adults 18-50 years)
Patient gender
Patient 5 digit zip code, primary residence
Race and ethnicity code or codes (up to 5) as defined in the MNCM REL Data Field Specifications and Codes
Country of origin as defined in the MNCM REL Data Field Specifications and Codes
Primary language as defined in the MNCM REL Data Field Specifications and Codes
Insurance coverage code as defined in the MNCM Insurance Coverage Data Field Specifications and Codes

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Risk adjustment model is estimated using a logistic model implemented in the SAS Procedure Glimmix that accounts for the measure´s non-continuous (binary) nature.

The dependent variable is Optimal Asthma Control. Risk factor variables include patient age, gender, insurance product, patient´s zip code, race/ethnicity and preferred language.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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 measure is calculated by submitting a file of individual patient values through a HIPAA secure data portal. Programming within the data portal determines if each patient is a numerator case and then a rate is calculated for each clinic site.

1)Is the patient´s DOB within the allowable time frame?
Yes>>Continue
No>>Patient not included in denominator
2)Has the patient had two office visits coded with an asthma diagnosis during the current and year prior to the measurement period?
Yes>>Continue
No>>Patient not included in denominator

3) Has the patient had one office visit for any reason during the measurement period?
Yes>> Patient included in denominator, continue
No>> Patient not included in denominator
4) Did the patient have an asthma control test within the measurement period?
Yes>> Continue
No>> Patient not included in numerator
5) Is the asthma control test tool used acceptable for the patient´s age?
Yes>> Continue
No>> Patient not included in numerator
6) Is the value of the control test equivalent to "in control"?
Yes>> Continue
No>> Patient not included in numerator
7) During the measurement period, was the patient asked about any hospitalizations or emergency department visits due to asthma in the 12 months prior?
Yes>>Continue
No>> Patient not included in numerator
8) Was the sum of patient reported emergency department visits and hospitalizations due to asthma in the prior 12 months equal to 0 or 1?
Yes>> Patient included in numerator
No>> Patient not included in numerator

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
MN Community Measurement encourages total population submission but also accepts sample submissions. The following are instructions for obtaining a sample:

Below are the requirements for submitting a sample:
>Each clinic must submit TWO samples – one for pediatric patients ages 5-17 (birthdate range includes MM/DD/YYYY to MM/DD/YYYY) and one for adult patients ages 18-50 (birthdate range includes MM/DD/YYYY to MM/DD/YYYY).
>If a clinic has less than 60 patients in either the pediatric or the adult population for the measure, submit ALL patients (e.g., if there are a total of 59 children ages 17 and under in the population for the measure, submit all 59 patients).
>If a clinic has 60 or more patients for each sample, first consider submitting all patients, otherwise you may submit a sample. The minimum required sample is 60 patients per age group, per clinic site, per measure (e.g., if there are 79 eligible patients in the population, first consider submitting all 79 patients, otherwise submit a sample of at least 60). MN Community Measurement recommends sampling the required 60 and adding an additional oversample of 20 patients to submit a total of 80 patients for each age group for each clinic.

To generate your data for the asthma measure, you will need to generate lists of random samples by age group and clinic. Each clinic needs one randomly sampled list of pediatric asthma patients and one randomly selected list of adult asthma patients.

SAMPLE STEP 1: CREATE SEPARATE RANDOMLY SELECTED LISTS FOR EACH AGE GROUP
a) First generate a list of ALL asthma patients ages 5-50 at a single clinic.
b) Break the list into two age groups:
>>Ages 5-17 in one list (birthdate range includes MM/DD/YYYY to MM/DD/YYYY)
>>Ages 18-50 in another list (birthdate range includes MM/DD/YYYY to MM/DD/YYYY)
c) Use one of the sampling methods for each age group (sampling methods are listed on the next page) to identify the patients in your denominator. This becomes the list of patients that you will need to look up the data for.
d) Repeat step “C” for the other age group.
e) Repeat steps a-d above for all of your clinics.

SAMPLE STEP 2: COMBINE YOUR PEDIATRIC AND ADULT SAMPLE BEFORE DATA SUBMISSION

After you generate all of your samples for all of your age groups for all clinic locations, you will need to combine the samples into one data file for your entire medical group. This one file will be the file you upload to the MN Community Measurement data portal.

NOTE: You will need to supply MNCM with patient counts BY AGE GROUP and BY CLINIC when you go to upload the data.

Example: You will be submitting only one file for your entire medical group and all the clinic sites within your organization. For example, if you have 4 family practice clinics, your final data file should have the following:

>Clinic Site 1 – 60 patients ages 5-17
>Clinic Site 1 – 60 patients ages 18-50
>Clinic Site 2 – 60 patients ages 5-17
>Clinic Site 2 – 60 patients ages 18-50
>Clinic Site 3 – 60 patients ages 5-17
>Clinic Site 3 – 60 patients ages 18-50
>Clinic Site 4 – 60 patients ages 5-17
>Clinic Site 4 – 60 patients ages 18-50

Sampling Methods

Method A: Excel Random Number Generator

For patient lists generated in Excel, use the "RAND" function to assign a random number to each record (please also see Microsoft Excel Help, topic RAND for more information):

1. Separate your list of patients into the two age groups and complete steps 2-10 for both your pediatric patients AND your adult patients
2. Insert a blank column on the leftmost side of the spreadsheet
3. Label new column "RAND"
4. Place cursor in the first blank cell (A2) and type =RAND()
5. Press enter (a number like 0.793958 will appear)
6. Place the cursor back into this cell; resting over the corner to have the pointer change to a black cross, double click or drag the formula down to the last row/patient
7. Highlight the whole column and click Edit, Copy, Paste Special = Values to freeze the random number (otherwise it will change with every click on the spreadsheet)
8. Sort entire patient population by this new random number
9. Work down the list row by row, starting with row 1 until the number of records in the sample is met for submission (at least 60 patients per clinic, plus at least 20 oversamples = 80 patients per clinic, per age group)
10. If a patient meets one of the accepted exclusions, keep working down the list and use oversamples that are after the number of records in the sample. For example, if 100 records will be submitted and 2 exclusions were found, include patient rows 101 and 102 to replace the excluded records.

Method B: Paper List Sample Selection

For paper-generated lists, complete the following steps:

1. Separate the list of patients into two age groups – do steps 2 and 3 below for both your pediatric patients AND your adult patients
2. Start with a list that has patients sorted by some unique patient related variable.
a. Identifying number like a medical record number [MRN] or chart number is ideal.
b. Sorting alphabetically is the least desirable in terms of randomness, however, this may be used when there is no other alternative.
3. Select every Nth patient for the number of patients that will be reported (at least 60 patients, plus at least 20 oversamples = 80 patients per clinic, per age group).
a. N should equal the clinic site's total population divided by the number of patients that will be submitted (if needed, round down to the nearest whole number). Review ALL randomly selected records and oversamples to exhaust the entire patient list. Highlight or mark every Nth patient on the list. This is the sample.
b. Example: If a clinic site has 800 asthma patients and 80 patients will be submitted, divide 800/80 = 10. Select every 10th patient on the list.

Proxy responses within PRO instruments are not allowed. The PRO´s must be completed in their entirety to be valid.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Capture of patient reported data is required for numerator compliance. Non-capture results in the patient being included in the denominator but counting as a numerator miss.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
All critical data elements are required for data submission. An incomplete PROM is invalid and is equivalent to an assessment not being completed. Patients that do not have a PROM for asthma control administered or documented during the measurement period are counted as a numerator miss. Patients for whom the number of ED and hospitalizations due to asthma are not documented are also counted as numerator noncompliant.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
An excel template with formatted columns for data fields is provided. Please refer to the attached data dictionary for data field definitions. All data is uploaded in electronic format (.csv file) to a HIPAA secure, encrypted and password protected data portal.

1. Asthma Control Test (ACT) and Childhood Asthma Control Test (C-ACT)
MNCM has secured permission for use of the ACT and C-ACT from GlaxoSmithKline for providers participating in quality measurement reporting to MNCM, under the following conditions:
• you will administer the instrument in a paper format only;
• permissible uses include only clinical care and quality measurement activities not related to research or publication;
• you may not modify the instrument or combine it with other instruments without prior written approval;
• the questions of the instrument must appear verbatim, in order, and together as they are presented and not divided on separate pages;
• for the ACT: the following trademark and copyright information must appear on the bottom of each page of the instrument and on all copies of the instrument; “Copyright 2002 by QualityMetric Incorporated. Asthma Control Test is a trademark of QualityMetric Incorporated.”
• for the C-ACT: the following acknowledgment be made as to the source and authorization for use of this material: “Copyright GSK. Used with permission.”
• you must utilize the instrument in its entirety;
• you agree to utilize only the most current version of the instrument as provided on MNCM’s Resource page.
• you agree to display the GSK logo as part of the instrument;
Of note, it IS permissible to record item responses and scores in an electronic health record, it IS NOT permissible to administer the instrument electronically to patients; i.e. kiosk, mobile device, patient portal.
2. Asthma Control Questionnaire (ACQ)
The ACQ is a copyrighted instrument available in various formats from the developer. Please visit the website http://www.qoltech.co.uk/acq.html for more information.
3. Asthma Therapy Assessment Questionnaire (ATAQ)
The ATAQ is copyrighted by Merck & Co., Inc, and available free of charge by going to: http://merckengage.qualitysolutionnavigator.com/ and navigating to the asthma resources. The Asthma Therapy Assessment Questionnaire (ATAQ) Adult should be used for patients 18 years and older. The Asthma Therapy Assessment Questionnaire (ATAQ) Pediatric should be used for patients 5 – 17 years old.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Optimal_Asthma_Control_Template_MeasSubm_CompositeMeasTesting_2015-09-01-635870704995158853.docx

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b7, 2d)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Composite Measure Title: Click here to enter measure title
Date of Submission: Click here to enter a date
Composite Construction:
☐Two or more individual performance measure scores combined into one score
☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

Instructions: Please contact NQF staff before you begin.

- If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.
- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all composite measures, sections 1, 2a2, 2b2, 2b3, 2b5, and 2d must be completed.**
- **For composites with outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitions** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2), validity (2b2-2b6), and composites (2d) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care

provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [15] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

2d. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2d1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2d2.the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(*if not conducted or results not adequate, justification must be submitted and accepted)*

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of

exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation. If different data sources are used for different components in the composite, indicate the component after the checkbox.*)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

1.3. What are the dates of the data used in testing? 7/1/2013 – 6/30/2014

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan

☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)
ADULTS: 415 clinics (All)
CHILDREN: 291 clinics (All)

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
ADULTS: 59,717 patients (All)
CHILDREN: 36,666 patients (All)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example: patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Race, ethnicity, language, country of origin, payer type (Commercial, Medicare, Medicaid, Self-Pay)

2a2. RELIABILITY TESTING
2a2.1. What level of reliability testing was conducted?
Note*: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. Describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

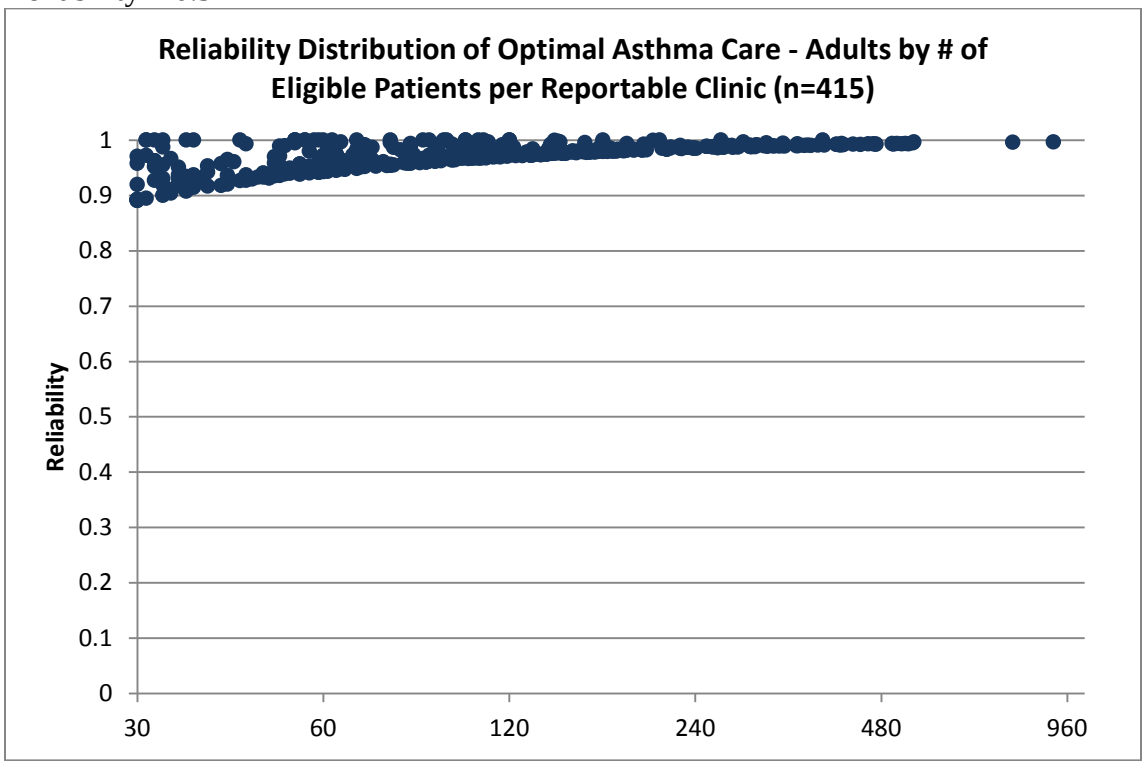
Used paper "Reliability in Provider Profiling" by John L. Adams, Ph.D as a reference
The BETABIN macro was used on each measure (SAS).

- First, we need to find the provider-to-provider variance:
  - $\sigma^2 = (\alpha\,\beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$
- Reliability = $\sigma^2 / (\sigma^2 + (p(1 - p)/n))$
  - p = rate
  - n = number of eligible patients
- Determine reliability rate for each provider.
- Average the reliability rate.

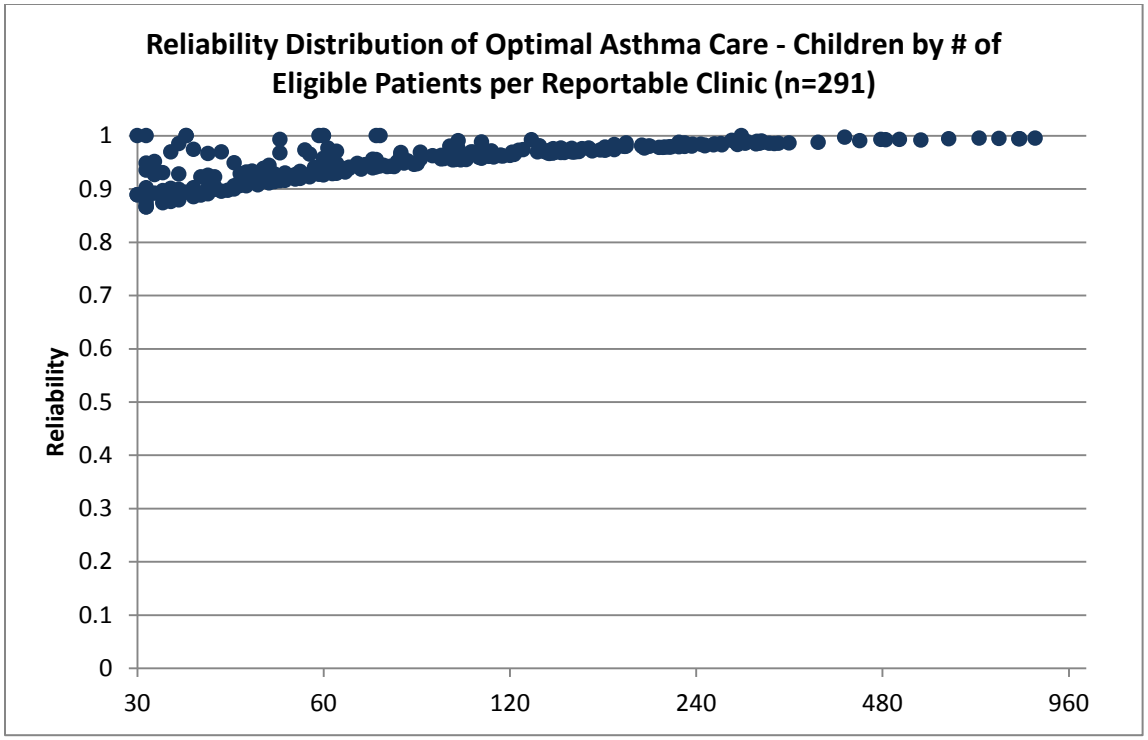
2a2.3. What were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

\

ADULTS:
Reliability = 0.972



CHILDREN:
Reliability = 0.951



2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
In terms of understanding reliability in detecting signal to noise, a reliability score of 0.70 or greater is considered acceptable for drawing conclusions about groups. This data analysis, along with precise specifications and excellent validation results of critical data elements, demonstrates this measure construct to be reliable and detect meaningful differences among provider groups.

2b2. VALIDITY TESTING
Note*: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*
2b2.1. What level of validity testing was conducted?

☒ **Composite performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

☐ **Systematic assessment of content validity**

☒ **Validity testing for component measures** *(check all that apply)*
Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
Composite performance measure score empirical validity testing:
Validity was tested for the computed composite score by testing the correlation of medical group performance with their performance on the Optimal Diabetes Care measure (NQF#0729). Asthma and diabetes are chronic conditions that require ongoing management of multiple risk factors in order to reduce a patient's overall risk. It is expected that the quality of care provided by a medical group to patient with diabetes would be of similar quality as the care provided to patients with asthma, and the respective performance measure scores should demonstrate such.

Validity testing for critical data elements:
Validating the submitted data via the direct data submission process is completed in four steps: denominator certification, data quality checks, validation audit, and the two-week medical group review period.

Denominator certification prior to data collection and extraction/ abstraction ensures that all medical groups apply the denominator criteria correctly and in a consistent manner. MNCM staff review the documentation to verify all criteria were applied correctly, prior to approval for data submission.
Denominator certification documentation for this measure includes:

- Date of Birth (ranges)
- Date of Service (ranges)
- ICD-9 Codes used
- Eligible specialties and provider types
- Exclusions to the measure and attest to mechanism for exclusions
- Attestations related to changes in medical record or billing systems
- Supplying all query code for review

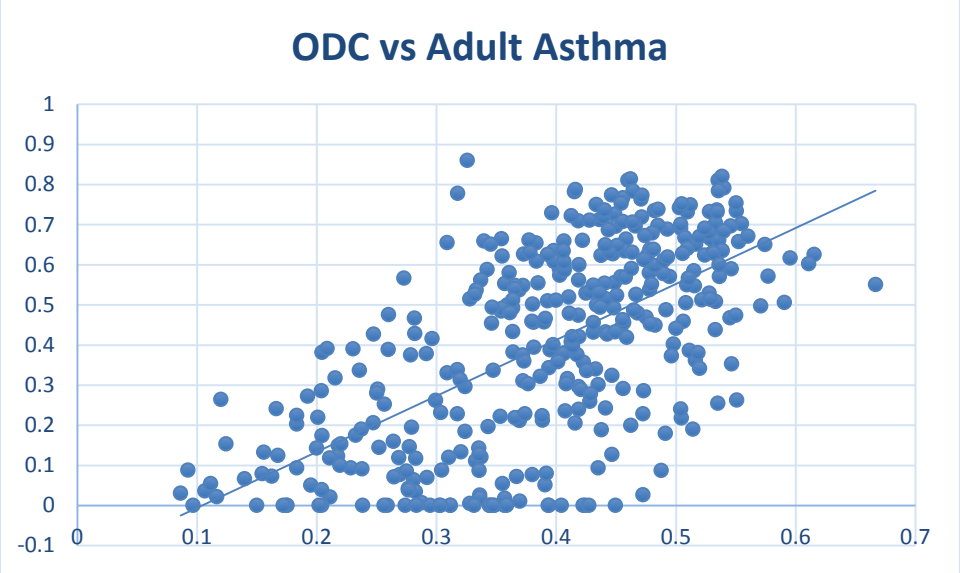
Common areas of correction in denominator for this measure included missing query code, incorrect date of birth ranges, incorrect dates for counting visits, missing ICD-9 codes or incomplete attestation. All were corrected prior to data submission.

Following data submission to the MNCM Data Portal, there are additional data quality checks in place for evaluating the accuracy of data submitted. During file upload, program checks for valid dates, codes and values and presents users with errors and warnings. Additionally, MNCM staff review population counts (denominator) and outcome rates for any significant variance from the previous year's submission and may prompt further clarification from the medical group.

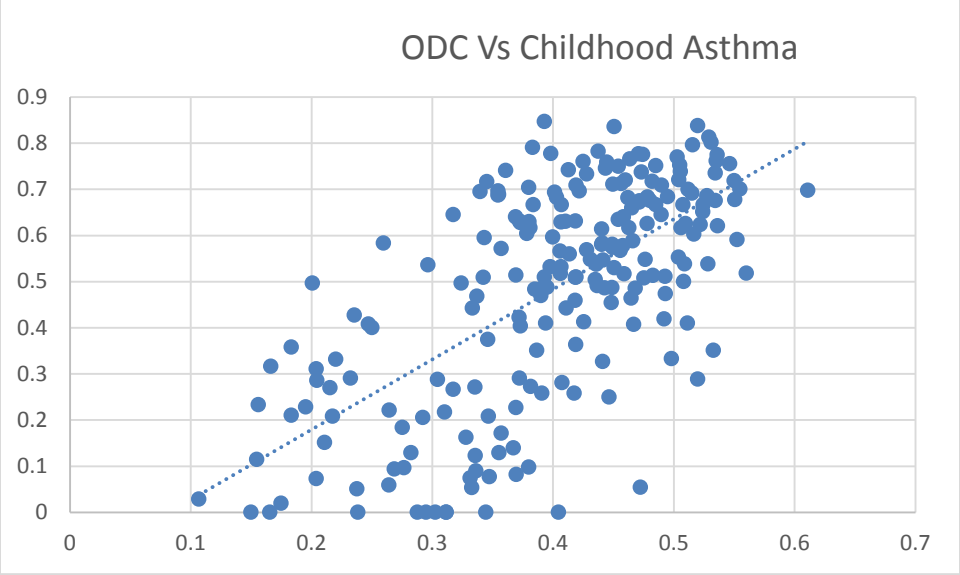
Validation audits verify that the clinical data submitted for the numerator component of the measure matched the data in the patient record. Other data elements are also audited to verify the patient was included in the denominator correctly.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Composite performance measure score empirical validity testing:
Based on linear regression analysis, a medical group's performance on the Optimal Diabetes Care measure is associated with its performance on the Optimal Asthma Control (Adults) measure, as demonstrated by a correlation coefficient of 0.63, representing a fairly strong correlation.



Based on linear regression analysis, a medical group's performance on the Optimal Diabetes Care measure is associated with its performance on the Optimal Asthma Control (Children) measure, as demonstrated by a correlation coefficient of 0.66, representing a fairly strong correlation.



Validity testing for critical data elements:
Initial validation audit results in 2014 demonstrated an 85% pass rate for the optimal asthma care measure, after which corrections were made and a second round of validation was conducted with the medical groups that failed to confirm accurate data submission. All but one medical group required to resubmit did so and passed subsequent audit. The remaining clinic chose not to participate in the audit process and was subsequently not included in reporting or aggregate calculations.

100% of groups achieved the desired > 90% data accuracy when submitted data was compared to medical record data (EMR or paper) of the patient.

Individual critical data element validation results (% agreement with true source):

Critical data elements	# of records with errors that do NOT alter result	% of records with errors that do NOT alter result	# of records with errors that alter result	% of records with errors that alter result	Total # audited
DOB	2	0.16%	0	0.00%	
Control test date	9	0.74%	1	0.08%	
Control test result	7	0.57%	0	0.00%	
ED/Hosp date	11	0.90%	7	0.57%	
ED #	5	0.41%	7	0.57%	
Hosp #	5	0.41%	7	0.57%	
Unique records	18	1.47%	8	0.66%	1221
Unique entities	5	4.27%	3	2.56%	117

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Composite performance measure score empirical validity testing:
As demonstrated by the correlation coefficient of 0.63 the total variation in performance on the Optimal Asthma Control measure (adults) can be explained by variation in the Optimal Diabetes Care measure. This degree of correlation indicates that the Optimal Asthma Control composite measure score accurately reflects the quality of care provided.
As demonstrated by the correlation coefficient of 0.66 the total variation in performance on the Optimal Asthma Control measure (children) can be explained by variation in the Optimal Diabetes Care measure. This degree of correlation indicates that the Optimal Asthma Control composite measure score accurately reflects the quality of care provided.

Validity testing for critical data elements:
High compliance with critical data element validity as demonstrated by annual validation audit processes.

2b3. EXCLUSIONS ANALYSIS
Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Required Exclusions	The following exclusions must be applied to the eligible population: ☐ Patient had a diagnosis of cystic fibrosis, COPD, emphysema or acute respiratory failure (*Obstructive Lung and Respiratory Failure* Value Set)

Allowable Exclusions	The following exclusions are allowed to be applied to the eligible population: ☐ Patient was a permanent nursing home resident at any time during the measurement period ☐ Patient was in hospice or receiving palliative care at any time during the measurement period ☐ Patient died prior to the end of the measurement period ☐ Documentation that diagnosis was coded in error

Clinics are asked to describe how they handle exclusions during the denominator certification process. Some clinics are able to automatically remove patients as part of their query for eligible patients (e.g., deceased patients or patients with an allowable exclusion diagnosis). Clinics also manually exclude patients during data abstraction.

MNCM requires clinics to track the individual patients that are excluded and submit the reasons for exclusion to MN Community Measurement.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
The volume of excluded patients is consistently less than 1% of the total population in both the adult and children populations.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
While the impact of the exclusions is small, the majority are optional exclusions where the practice can choose whether or not to apply the exclusions. The required exclusions are able to be programmed into query and represent a patient population that is clinically different than the target population.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used? *(check all that apply)*
☐ **Endorsed (or submitted) as individual performance measures**
☐ **No risk adjustment or stratification**
☒ **Statistical risk model**
☐ **Stratification by risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

During our measure development process, the expert panel discusses potential variables for risk adjustment that are important to consider for the measured population. Variables are included in public comment and collected during pilot testing to assess feasibility. For this measure, the measure development workgroup identified comorbidity of depression, tobacco use and tobacco exposure as data elements that should be collected and study. These factors are in addition to the standard demographic variables (gender, age, zip, race/ethnicity, country of origin, primary language and insurance product). The potential risk adjustment variables are then evaluated for appropriate inclusion in the model based on a t value outside the range of -2.0 and +2.0.

The dependent variable was optimal asthma control (OAC). Measures included for risk adjustment that were submitted by clinics included patient age, gender, insurance product, and patient's zip code. For the analysis, MSHO, Medicaid, Uninsured/Self-Pay, and Special Needs were grouped into one insurance product category. For analysis purposes, patient age was categorized as 18 to 25 and 26 to 50. Indicators were included for female gender, smoking status, and exposure to secondhand smoke. The latitude and longitude of the centroid associated with the patient and clinic zip codes were used to calculate the distance from the patient zip code to the clinic zip code. The patient zip code was linked to zip code data from the Primary Care Service Area project to obtain measures of median income, percent African American, and percent Hispanic. These serve as proxies for direct measures of income and race/ethnicity (African American and Hispanic).

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The effect of risk adjustment on clinic ranking is examined in three ways. First, clinic unadjusted and adjusted quality measures are compared using correlation analysis. Two types of correlation are used, Pearson and Kendall. Pearson's correlation examines the correlation when the measures are treated as continuous measures. A high correlation (close to 1) means that the two measures strongly co-vary; when one is high the other is high. Kendall's correlation examines the similarity between the unadjusted and adjusted quality measure in terms of the similarity in the way clinics are ranked by the measures. Because of the focus of Kendall's correlation on comparing ranks and the interest in the use of clinic quality scores for clinic comparison, Kendall's correlation is likely to be the most useful correlation measure.

The second comparison ranks the clinics into performance rank deciles based on the unadjusted and adjusted scores and then examines how decile rankings based on unadjusted measures compare to decile rankings based on adjusted measures.

The third comparison ranks clinics into categorizing clinics into Poor, Below Average, Average, Above Average, and Excellent using statistical methods that take into account the quality measure's confidence interval which is calculated based on the number of patients each clinic reports(7, 8). These two methods are compared directly in our accompanying report on the quality deviations ranking approach.

At the patient level, the average OAC was 26.2% (standard deviation = 23.6). The average number of patients reported by a clinic was 139 (standard deviation = 121). The average age in the examined population was 35.2 years. Within the population, 66% were female, 68.9% had commercial insurance, 5.72% had Medicare coverage, and 15.5% had Medicaid coverage.

Risk adjustment is necessary only when there is heterogeneity across clinics. There was significant heterogeneity across clinics in insurance product mix ($\chi2 = 13{,}309$, $p < .001$), patient age ($\chi2 = 2{,}798$, $p < .001$), gender ($\chi2 = 1035$, $p < .001$), tobacco use ($\chi2 = 82539$, $p < .001$), second hand tobacco exposure ($\chi2 = 95100$, $p < .001$) and distance to the clinic ($\chi2 = 27{,}061$, $p < .001$).

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

The T value approximately shows how many standard deviations the estimated coefficient is from zero. Typically, T values of less than -2 or greater than 2 are significant at the 95% confidence level (less than 1 chance in 20 of the estimated effect occurring by chance). The odds ratio shows the odds of the risk adjuster affecting OAC when compared with the contrast (e.g., the effect of distance to a clinic of greater than 20 miles is compared with the contrast category of less than 5 miles). The upper and lower bounds of the 95% confidence interval indicate the interval within which there is strong statistical confidence the true odds ratio lies. There is less than 1 chance in 20 of the true odds ratio being outside the 95% confidence interval. When a confidence interval includes the value 1 the effect is interpreted as not statistically significant.

The effects of all of the risk adjusters except insurance product are not statistically significant, meaning that these variables are related to quality measurement. In comparison to patients with commercial insurance, patients with all other types of insurance are less likely to have optimal diabetes care. Patients who are MSHO, Medicaid, Special Needs, Uninsured or self-pay are 37% less likely than patients with commercial insurance to have optimal asthma care.

These results imply that clinics with a higher proportion of patients not having commercial insurance experience adverse selection.

TABLE 1: EFFECT OF POTENTIAL RISK ADJUSTERS ON OAC						
Variable	Contrast	Estimate	T-value	Odds Ratio	Lower 95% CI	Upper 95% CI
1A: MODEL WITHOUT SES AND RACE FROM ZIP CODE DATA						
Age						
26-50	18-25	-0.05	-1.59	0.95	0.90	1.01
Gender						
Female	Male	-0.01	-0.52	0.99	0.94	1.04
Distance from Clinic						
<5 miles	Same Zip	0.03	0.91	1.04	0.96	1.11
5-10 miles	Same Zip	0.01	0.21	1.01	0.94	1.08
10-20 miles	Same Zip	0.00	-0.05	1.00	0.92	1.08
20+ miles	Same Zip	-0.08	-1.51	0.92	0.84	1.02
Insurance						
Medicare	Commercial	-0.47**	-8.56	0.63**	0.56	0.70
Medicaid / MSHO / Special Needs / Self-pay / Uninsured	Commercial	-0.58**	-18.74	0.63**	0.56	0.70
Constant		-1.19**	-2.87			
1B: MODEL WITH SES AND RACE FROM ZIP CODE DATA						
Age						
26-50	18-25	-0.04	-1.50			
Gender						
Female	Male	-0.01	-0.49			
Distance from Clinic						
<5 miles	Same Zip	0.04	1.08			
5-10 miles	Same Zip	0.01	0.21			
10-20 miles	Same Zip	-0.01	-0.20			
20+ miles	Same Zip	-0.07	-1.40			
Insurance						
Medicare	Commercial	-0.45**	-8.28			
Medicaid / MSHO / Special Needs / Self-pay / Uninsured	Commercial	-0.56**	-18.08			
Zip Code Data						
Median Income		0.03*	2.52			
Percent Black		0.00	-1.48			
Percent Hispanic		0.18	0.56			
Constant		-1.31**	-3.14			
** indicates statistical significance.						

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Because OAC is a binary variable (0 or 1), the risk adjustment model was estimated using a logistic model implemented in the SAS Procedure Glimmix that accounts for its non-continuous nature. The risk adjusters and an indicator for each clinic were included in the model. The estimated coefficient for the clinic indicator measures the clinic's OAC adjusting for the patient risk adjusters that were included in the model. The clinic level indicator was used to construct a risk adjusted OAC score at the clinic level that ranged from 0 to 1 (0% to 100%). The effect of risk adjustment on clinic rankings was estimated by comparing the risk adjusted OAC to the unadjusted OAC measure, the average OAC for all patients submitted by the clinic. The risk adjustment for tables 2-4 includes all risk adjustment variables detailed in Table 1 (age, gender, distance, and insurance). Since age greater than 65, the contrast category for age effects, captures the effect of age, the Medicare indicator for insurance captures the effect of Medicare independent of age.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
We tested the overall correlation between the unadjusted and risk adjusted OAC measure using two methods, a Pearson correlation and a Kendall's Tau correlation. In both cases, the value 1 represents a perfect correlation and the value 0 represents a complete lack of correlation between unadjusted and adjusted measures. The Pearson correlation compares the risk adjusted and unadjusted clinic OAC values, and is 0.99 which shows a very strong correlation between the unadjusted and adjusted OAC measure. The Kendall's Tau correlation compares unadjusted and adjusted rank order of clinics, and was 0.93. This is still a strong correlation, but not as strong as the .99 correlation between risk adjusted and unadjusted clinic values.
We used various methods to compare the effect of risk adjustment on clinic rank (risk adjustment includes all variables detailed in Table 1), as shown in Tables 2 through 4. Table 2
compares the unadjusted and adjusted decile ranking of clinics (the decile approach). Table 3 compares unadjusted and adjusted clinic quality rankings based on their statistical difference from the OAC population mean (the quality deviations approach). Table 4 compares the adjusted clinic rankings between both the decile and quality deviations approaches.
Our analysis of the decile approach shows that, consistent with the Kendall's Tau correlation analysis, there are not major differences between the adjusted and unadjusted clinic rankings by decile (shown in Table 2). Most clinics (272) remain in the diagonal, which indicates no change in clinic ranking due to risk adjustment, while some (29) increase in ranking and others (63) decrease in ranking. Table 3 compares the unadjusted and adjusted clinic rankings using the quality deviations approach. Consistent with the decile ranking approach, Table 3 shows that the majority of clinics (315) experience no change in clinic ranking due to risk adjustment while a few (20) increase in rank and a few (29) decrease in rank.

Table 2 – Comparison of Unadjusted and Adjusted Decile Ranks (N / Percent of Row)*

Unadjusted Decile Rank	Risk Adjusted Decile Rank										
	0 to 10%	10% to 20%	20% to 30%	30% to 40%	40% to 50%	50% to 60%	60% to 70%	70% to 80%	80% to 90%	90% to 100%	Total
0% to 20%	35 48.61	35 48.61	2 2.78	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	72
20% to 30%	1 2.70	1 2.70	32 86.49	3 8.11	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	37
30% to 40%	0 0.00	0 0.00	3 8.33	32 88.89	1 2.78	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	36
40% to 50%	0 0.00	0 0.00	0 0.00	1 2.70	33 89.19	3 8.11	0 0.00	0 0.00	0 0.00	0 0.00	37
50% to 60%	0 0.00	0 0.00	0 0.00	0 0.00	3 8.33	30 83.33	3 8.33	0 0.00	0 0.00	0 0.00	36
60% to 70%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	3 8.11	27 72.97	7 18.92	0 0.00	0 0.00	37
70% to 80%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	7 18.92	23 62.16	7 18.92	0 0.00	37
80% to 90%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	6 16.67	28 77.78	2 5.56	36
90% to 100%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	2 5.56	34 94.44	36
Total	36	36	37	36	37	36	37	36	37	36	364

*Grey cells indicate no change in rank. Blue cells indicate increase in rank after risk adjustment, and red cells indicate decrease in rank after risk adjustment. N is the number of clinics in each cell, and the percent of row is the percent of the total unadjusted decile ranked clinics in each cell.

**Some rows missing due to unusual distribution with too many zero or low scoring clinics makes it impossible to split them into separate deciles.

Table 3 – Comparison of Unadjusted and Adjusted Deviations Quality Rankings (N / Percent of Row)

Unadjusted Deviations Quality	Risk Adjusted Deviations Quality					
	Poor (3+ SD below mean)	Below Average (2-3 SD below mean)	Average (mean + or – 2 SD)	Above Average (2-3 SD above mean)	Excellent (3+ SD above mean)	Total
Poor (3+ SD below mean)	138 95.17	7 4.83	0 0.00	0 0.00	0 0.00	145
Below Average (2-3 SD below mean)	2 13.33	12 80.00	1 6.67	0 0.00	0 0.00	15
Average (mean + or – 2 SD)	0 0.00	4 6.15	60 92.31	1 1.54	0 0.00	65
Above Average (2-3 SD above mean)	0 0.00	0 0.00	2 15.38	8 61.54	3 23.08	13
Excellent (3+ SD above mean)	0 0.00	0 0.00	0 0.00	14 11.11	112 88.89	126
Total	140	23	63	23	115	364

* Grey cells indicate no change in rank. Blue cells indicate increase in rank after risk adjustment, and red cells indicate decrease in rank after risk adjustment. N is the number of clinics in each cell, and the percent of row is the percent of the total unadjusted decile ranked clinics in each cell.

Table 4 – Comparison of Risk Adjusted Decile and Deviations Quality Rankings (N / Percent of Row)						
Risk Adjusted Decile	Risk Adjusted Deviations Quality					
	Poor (3+ SD below mean)	Below Average (2-3 SD below mean)	Average (mean + or – 2 SD)	Above Average (2-3 SD above mean)	Excellent (3+ SD above mean)	Total
0 to 10%	36 100.00	0 0.00	0 0.00	0 0.00	0 0.00	36
10% to 20%	36 100.00	0 0.00	0 0.00	0 0.00	0 0.00	36
20% to 30%	35 94.59	2 5.41	0 0.00	0 0.00	0 0.00	37
30% to 40%	26 72.22	10 27.78	0 0.00	0 0.00	0 0.00	36
40% to 50%	7 18.92	11 29.73	19 51.35	0 0.00	0 0.00	37
50% to 60%	0 0.00	0 0.00	36 100.00	0 0.00	0 0.00	36
60% to 70%	0 0.00	0 0.00	8 21.62	16 43.24	13 35.14	37
70% to 80%	0 0.00	0 0.00	0 0.00	7 19.44	29 80.56	36
80% to 90%	0 0.00	0 0.00	0 0.00	0 0.00	37 100.00	37
90% to 100%	0 0.00	0 0.00	0 0.00	0 0.00	36 100.00	36
Total	140	23	63	23	115	364

*N is the number of clinics in each cell, and the percent of row is the percent of the total unadjusted decile ranked clinics in each cell.

Table 5 – Comparison of Unadjusted and Product Adjusted Decile Ranks (N / Percent of Row)*

Unadjusted Decile Rank	Risk Adjusted Decile Rank										
	0 to 10%	10% to 20%	20% to 30%	30% to 40%	40% to 50%	50% to 60%	60% to 70%	70% to 80%	80% to 90%	90% to 100%	Total
0% to 20%	35 48.61	35 48.61	2 2.78	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	72
20% to 30%	1 2.70	1 2.70	32 86.49	3 8.11	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	37
30% to 40%	0 0.00	0 0.00	3 8.33	31 86.11	2 5.56	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	36
40% to 50%	0 0.00	0 0.00	0 0.00	2 5.41	32 86.49	3 8.11	0 0.00	0 0.00	0 0.00	0 0.00	37
50% to 60%	0 0.00	0 0.00	0 0.00	0 0.00	3 8.33	30 83.33	3 8.33	0 0.00	0 0.00	0 0.00	36
60% to 70%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	3 8.11	28 75.68	6 16.22	0 0.00	0 0.00	37
70% to 80%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	6 16.22	24 64.86	7 18.92	0 0.00	37
80% to 90%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	6 16.67	29 80.56	1 2.78	36
90% to 100%	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	1 2.78	35 97.22	36
Total	36	36	37	36	37	36	37	36	37	36	364

*Grey cells indicate no change in rank. Blue cells indicate increase in rank after risk adjustment, and red cells indicate decrease in rank after risk adjustment. N is the number of clinics in each cell, and the percent of row is the percent of the total unadjusted decile ranked clinics in each cell.
**Some rows missing due to unusual distribution with too many zero or low scoring clinics makes it impossible to split them into separate deciles.

Table 6 – Comparison of Unadjusted and Product Adjusted Deviations Quality Rankings (N / Percent of Row)

Unadjusted Deviations Quality	Risk Adjusted Deviations Quality					
	Poor (3+ SD below mean)	Below Average (2-3 SD below mean)	Average (mean + or – 2 SD)	Above Average (2-3 SD above mean)	Excellent (3+ SD above mean)	Total
Poor (3+ SD below mean)	138 95.17	7 4.83	0 0.00	0 0.00	0 0.00	145
Below Average (2-3 SD below mean)	2 13.33	12 80.00	1 6.67	0 0.00	0 0.00	15
Average (mean + or – 2 SD)	0 0.00	5 7.69	58 89.23	2 3.08	0 0.00	65
Above Average (2-3 SD above mean)	0 0.00	0 0.00	2 15.38	9 69.23	2 15.38	13
Excellent (3+ SD above mean)	0 0.00	0 0.00	0 0.00	12 9.52	114 90.48	126
Total	140	24	61	23	116	364

* Grey cells indicate no change in rank. Blue cells indicate increase in rank after risk adjustment, and red cells indicate decrease in rank after risk adjustment. N is the number of clinics in each cell, and the percent of row is the percent of the total unadjusted decile ranked clinics in each cell.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
See above

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
see above

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Our analysis of risk adjustment factors for the Optimal Asthma Control measure indicates that insurance provider variables are related and may warrant attention for risk adjustment.

For most clinics there is no change in clinic ranking due to risk adjustment, while some increase in ranking and others decrease in ranking. For those whose rankings are impacted by the risk adjustment, it is legitimate and based on disparate differences among these clinics.

***2b4.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
Note: *Applies to the composite performance measure.*

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
Not all patients have the same likelihood of achieving optimal health outcomes due to barriers based on demographic, physical or socio-economic situations. Risk Adjustment is the process of adjusting the measure to account for the barriers that are outside the control or influence of the provider. Appropriate risk adjustment creates a more normal distribution of results.

The risk adjustment is done using an Actual to Expected methodology. This methodology does not alter the result of a clinic/medical group; the actual rate remains unchanged but instead of comparing the rate to the raw market average, a unique expected rate based on the proportion of each risk category for each clinic/medical group is used for comparison.

The comparison between clinics is the aforementioned test of significance and the actual to expected ratio: actual percentage of patients meeting criteria divided by the expected percentage of patients meeting criteria for the particular entities mix of patient risk.

To test whether or not there was a statistically significant difference between the expected rate and the actual rate achieved by a clinic/medical group, a one population proportions test was used. This method is employed to test the proportion of optimally managed patients attributed to a clinic compared to a specific value for that clinic. In the MNCM case, the specified value is an expected rate calculated taking into account the overall state rate and adjusted for risk factors specific to the measure. The methodology uses a 99% test of significance.

Patients from each clinic/medical group were categorized into one of the four major insurance types (Commercial; Medicare; MN Government Programs; and Self-Pay/Uninsured). Market average rates (percent of patients meeting the measure) were calculated for each risk category. The expected rate is then calculated for each clinic/medical group using the proportions of each category.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Asthma – Adult

Risk Variables: Product

Clinic Distribution 2014 Dates

	Below	Expected	Above	Total
Below	131	14	0	145
Average	0	110	5	115
Above	0	11	117	128
	131	135	122	388

Better	19	4.9%
Same	358	92.3%
Worse	11	2.8%

Asthma – Child

Risk Variables: Product

Clinic Distribution 2014 Dates

	Below	Expected	Above	Total
Below	71	21	0	92
Average	0	95	7	102
Above	0	4	73	77
	71	120	80	271

Better	28	10.3%
Same	239	88.2%
Worse	4	1.5%

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

More than 50% of groups who submit data to MNCM have results that pass the test for significance to be meaningfully different than the expected rate for their patient population.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

If only one set of specifications for each component, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Note: *Applies to the overall composite measure.*

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

For this patient level all-or-none composite measure, elements missing from any component (e.g. visit but no blood pressure during the measurement year) are counted as a numerator component fail and therefore the patient would be accounted for and remain in the denominator.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for*

handling missing data that were considered and pros and cons of each)
For this measure, missing data is a representation that the patient was not assessed during the required timeframe, and therefore represents an absence of activity, not an inability to gather the data field from the record. Missing data is not a factor for this measure.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH
Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*
This composite measure is a patient level all-or-none composite in which the desired goal is for the patient to demonstrate both short term control (asthma control patient reported outcome) and long-term control (less than 2 ED visits and hospitalizations due to asthma in the past 12 months). Optimal Asthma Control is which both short AND long term control are achieved.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)
Additional components considered included inclusion of an asthma management plan and avoidance of tobacco exposure.
An asthma management plan was considered because of the strong body of evidence that participation in a comprehensive self-management and education program delivery leads to improved outcomes, however, the most feasible proxy for assessing this was the presence of a written asthma action plan in the record, which does not have a strong and consistent body of evidentiary support for leading to improved outcomes. This component was removed from consideration due to the inconsistent evidentiary support.
Exposure to tobacco is a known trigger for asthmatic exacerbations, however, due to the physician's limited ability to influence the behaviors of individuals in the home of an asthma patient, this component was removed from consideration.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*
The patient reported outcome component demonstrates short term control (7 day to 4 week recall).
The number of ED visits and hospitalization due to asthma in the prior 12 months demonstrates long term control.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.e., *what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
MN Community Measurement has modeled the direct data submission process to minimize inaccuracies, errors, and unintended consequences. All groups participating sign a terms of use agreement that delineates the group´s responsibilities for submission of data and consequences for not participating in good faith. Additionally, all groups sign a Business Associate Agreement that outlines the use of the data.

The denominator certification process prior to any data collection ensures that groups are following the specifications and correctly identifying their population and serves as a point of correction prior to the expenditure of resources for data collection. Groups provide documentation of cases that are excluded and this is reviewed by MN Community Measurement staff prior to approval of the data submission. Extensive audit processes also support the data´s accuracy. After data submission, onsite validation audits are conducted comparing the submission to the patient´s medical record using NCQA´s 8 and 30 rule for audit requiring a 90% accuracy rate. Audits are conducted for the all clinic locations during their first data submission.
1. Specifications- Detailed specifications with instructions on how to handle most data collection situations has been valuable to medical groups and provides increased data accuracy. MN Community Measurement convened a technical workgroup to discuss enhancements for the next version of the data collection guide.
2. Audit- Audit methods have ensured the accuracy of our data, thus, clinics can be compared successfully because everyone pulls data the same way and are subject to the same rules.
3. Confidentiality- MN Community Measurement only receives the patient level information needed to calculate the rates, determine eligibility for inclusion in the measure and support the administration of pay for performance programs. The PHI submitted is minimal and the data is protected by 1) password protection with password only available to the medical group submitting data, 2) file upload process is encrypted as data is transferred and 3) data is stored on a separate secure server and meets all HIPAA protection rules.
4. Electronic Medical Record- It is easier for groups that have an electronic medical record to submit their full population of patients, however groups with paper chart systems can successfully submit a sample if necessary.
5. Data Collection Burden - The submission timeframe is a mid-year cycle (July 1 – June 30) versus a calendar year cycle (January – December). This was planned to reduce data collection burden during the first part of the year. Most groups were able to meet the submission deadline (6 weeks after the end of the measurement period), however MN Community Measurement allowed other groups to submit data after the submission deadline since it was the first year the data was reported.
6. State Requirements & Health Plans - State mandated reporting and pay for performance programs impacts the number of groups that submit data.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None by MN Community Measurement.
MN Community Measurement has obtained permissions for use for the ACT and Childhood ACT for practices who participate in quality reporting to MNCM.
Practices that wish to use the ACQ, ATAQ or ACT beyond the current agreement must obtain permissions for use from the tool licensor.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting MNHealthScores http://www.mnhealthscores.org/asthma-children-842 MNHealthScores http://www.mnhealthscores.org/medical-group-measure-detail/adults/#/results Payment Program PQRS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/pqri/ Regulatory and Accreditation Programs MN Health Care Homes http://www.health.state.mn.us/healthreform/homes/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) MN Health Care Homes http://www.health.state.mn.us/healthreform/homes/

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

MNHealthScores is the consumer facing website hosted by MN Community Measurement where visitors can find quality, cost and patient experience information to aid health care decision making. It has information covering the entire state of MN and the surrounding border communities.

MN Health Care Homes is a certification program ran by the Minnesota Department of Health.

This measure is also used by multiple health plans in the market in provider contracting.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

This measure was revised in Report Year 2015 and does not have comparable historical data to demonstrate progress on

improvement.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

No

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications completely harmonized?

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

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** MN Community Measurement
Co.2 **Point of Contact:** Jasmine, Larson, jlarson@mncm.org, 612-746-4514-
Co.3 **Measure Developer if different from Measure Steward:** MN Community Measurement
Co.4 **Point of Contact:** Jasmine, Larson, jlarson@mncm.org, 612-746-4514-

Additional Information

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

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2010
Ad.3 **Month and Year of most recent revision:** 03, 2015
Ad.4 **What is your frequency for review/update of this measure?** Annual
Ad.5 **When is the next scheduled review/update for this measure?** 03, 2016

Ad.6 **Copyright statement:** © MN Community Measurement, 2015. All rights reserved
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2856
De.2. **Measure Title:** Pharmacotherapy Management of COPD Exacerbation
Co.1.1. **Measure Steward:** National Committee for Quality Assurance
De.3. **Brief Description of Measure:** This measure assesses the percentage of COPD exacerbations for patients 40 years of age and older who had an acute inpatient discharge or ED encounter on or between January 1–November 30 of the measurement year and who were dispensed appropriate medications.

Two rates are reported.
1. Dispensed a systemic corticosteroid (or there was evidence of an active prescription) within 14 days of the event
2. Dispensed a bronchodilator (or there was evidence of an active prescription) within 30 days of the event

Note: The eligible population for this measure is based on acute inpatient discharges and ED visits, not on patients. It is possible for the denominator to include multiple events for the same individual.
1b.1. **Developer Rationale:** This measure assesses whether patients who had a hospitalization or an emergency department (ED) visit for a COPD exacerbation were provided appropriate medication (systemic corticosteroids and bronchodilators) to treat symptoms and prevent future exacerbations. The improvement in quality envisioned by the use of this measure is to increase the use of systemic corticosteroids and bronchodilators following a COPD exacerbation in order to shorten patients' recovery time, improve lung function and reduce the risk of early relapse, treatment failure, and length of hospital stay.

S.4. **Numerator Statement:** Numerator 1 (Systemic Corticosteroids): The number of patients dispensed a prescription for systemic corticosteroid on or 14 days after the Episode Date*. Count systemic corticosteroids that are active on the relevant date.

Numerator 2 (Bronchodilator): The number of patients dispensed a prescription for a bronchodilator on or 30 days after the Episode Date*. Count bronchodilators that are active on the relevant date.

*The Episode Date is the date of service for any acute inpatient discharge or ED claim/encounter during the 11-month intake period with a principal diagnosis of COPD.
S.7. **Denominator Statement:** All patients age 40 years or older as of January 1 of the measurement year with a COPD exacerbation as indicated by an acute inpatient discharge or ED encounter with a principal diagnosis of COPD.
S.10. **Denominator Exclusions:** 1) Exclude episode dates when the patient was transferred directly to an acute or nonacute inpatient care setting for any diagnosis.
2) Exclude episode dates when the patient was readmitted to an acute or nonacute inpatient care setting for any diagnosis within 14 days after the episode date.
3) Exclude episode dates when the patient had an ED visit for any diagnosis within 14 days after the Episode date.

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Health Plan, Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

This measure was previously endorsed as NQF #0549, but endorsement was removed in the last pulmonary project (July 2012); the previous Committee encouraged the developer to re-submit the measure at the next opportunity. The specifications for this measure, #2856, appear unchanged and so information related to #0549 also is provided for context and completeness.

The developer provides the following information:

- The level of analysis is Health Plan/Integrated Delivery System.
- This measure assesses whether patients who were hospitalized or had an ED visit for a COPD exacerbation were dispensed appropriate medication to treat their symptoms; the measure is event-based, not patient-based. Evidence of an active description on the day of event also counts.
- Systemic corticosteroids and bronchodilators following a COPD exacerbation improve symptoms reduce the risk of early relapse and shorten recovery time and length of hospital stay.
- Evidence is based on 2 different clinical practice guidelines:
    1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Dated 2015 and encompassing 613 studies from 1965-2015, the body of evidence includes the highest level, Evidence A. The guideline did not provide a breakdown of the specific number of RCTs. The recommendation related to systemic corticosteroid and short-acting bronchodilators to treat COPD exacerbations was based on 25 studies, including RCTs, meta-analyses, systematic reviews of RCTs, and observational studies.
    2. Institute for Clinical Systems Improvement (ICSI). Dated 2013, the ICSI guidelines cover a body of evidence covering 1968-2012. The ICSI guideline recommendations related to short-acting bronchodilators following an exacerbation referenced one meta-analysis of 12 RCTs. The ICSI guideline recommendations related to systemic corticosteroids following an exacerbation referenced 4 RCTs and one observational study.
- Per the measure developer, the overall quality of evidence that systemic corticosteroids improve outcomes following COPD exacerbations is high and based on multiple RCTs conducted with hundreds of patients with COPD; these studies consistently found patients who received systemic corticosteroids within 2 weeks following an exacerbation had improved lung function, fewer treatment failures, increased time to relapse, and shorter hospital stays than patients who did not receive systemic corticosteroids.
- Per the measure developer, the overall quality of evidence that short-acting bronchodilators improve outcome following COPD exacerbations is moderate because the RCTs were conducted in fewer patients, however findings from the meta-analyses found short-acting bronchodilators caused statistically significant increases in pulmonary function tests and are important for rapid improvement in shortness of breath and wheezing.
- Per the NQF Algorithm, the evidence is based on systemic reviews that include grading, and the measure developer provided an articulation of the overall quality, quantity, and consistency of the evidence.

Exception to evidence: Not applicable

Guidance from the Evidence Algorithm: 1->3->4->5 (highest eligible rating is HIGH)

Question for the Committee:

- *The measure has two numerators: 1) dispensed a systemic corticosteroid (or evidence of an active prescription within 14 days of the event; and 2) dispensed a bronchodilator (or evidence of an active prescription) within 30 days of the event. Is the evidence directly applicable to each individual process of care being measured?*

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following:

- The National Heart, Lung, and Blood Institute (NHLBI) states that more than 15 million adults have been diagnosed with chronic obstructive pulmonary disease (COPD), and that the actual number of those with the disease may be higher (NHLBI 2013). COPD mortality has risen, making it the third leading cause of death in the U.S. (Hoyert and Xu 2012; NHLBI 2013). Without intervention, deaths from COPD are projected to increase by more than 30 percent in the next 10 years, and COPD is projected to be the third leading cause of death worldwide (WHO 2014). In 2010, COPD was the main cause for 10.3 million physician visits, 1.5 million emergency department visits and 699,000 discharges (Ford et al, 2013).
- Performance data for the measure were extracted from the NCQA Healthcare Effectiveness Data and Information Set (HEDIS) for 2012, 2013, and 2014 (commercial HMO and PPOs; Medicare; Medicaid)
  - The mean systemic corticosteroid rate ranged from 65%-76% among various plans with little change seen from 2012-2014 within each plan type (</=2%).
  - The mean bronchodilator rate ranged from 78%-81% among various plans with little change seen from 2012-2014 within each plan type (</=2%).
  - In 2014, for the Systemic Corticosteroid Indicator, the developer reports:
    - A 16 percentage difference between commercial HMO plans in the 10th percentile and those in the 90th percentile
    - An 18 percentage difference between commercial PPO plans in the 10th percentile and those in the 90th percentile
    - A 30 percentage difference between Medicaid plans in the 10th and 90th percentiles.
  - In 2014, for the Bronchodilator Indicator, the developer reports:
    - A 16 percentage difference between commercial HMO plans in the 10th percentile and those in the 90th percentile
    - A 17 percentage difference between commercial PPO plans in the 10th percentile and those in the 90th percentile
    - A 25 percentage difference between Medicaid plans in the 10th and 90th percentiles

Disparities

- NCQA does not currently collect performance data stratified by race, ethnicity, or language.
- Medicaid vs. commercial performance is one proxy for disparities. For 2014 HEDIS data:
  1. Systemic Corticosteroids Indicator:
     - Mean Commercial HMO rate= 75% with <2% change
     - Mean Commercial PPO rate= 73% with <2% change
     - Mean Medicaid rate (HMO and PPO)= 66% with <1% change
  2. Bronchodilator Indicator:
     - Mean Commercial HMO Rate= 81% with <1% change
     - Mean Commercial PPO rate= 78% with <1 % change
     - Mean Medicaid rate (HMO and PPO)=81% with <2% change

- The developer provides a summary of literature that addresses disparities in care in this area:
  - Healthy People 2020 data show that, since 2007, White adults over age 45 have had a higher mortality rate than other races and ethnicities, while African American adults over age 45 have experienced higher hospitalization and higher emergency department (ED) utilization (U.S. Department of Health and

Human Services 2013).

- One study found that while there was variation in utilization, there was little variation in short-term outcomes following acute exacerbation. Predictors of increased ED visits were male sex, African American, Medicaid insurance, low socioeconomic status, and certain comorbidities; patients with high numbers of ED visits had higher frequency of a repeat admission within 30 days. When in the ED, African American and Hispanic patients were less likely to receive certain services, such as chest radiography, but were more likely to receive systemic corticosteroids both in the ED and at discharge; however, other disease and symptom management in the ED was not found to be significantly different between races. After adjusting for patient and ED characteristics, quality of care was not found to differ significantly between different races and ethnicities (Tsai et al. 2009).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *What does the general lack of improvement over time indicate?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments:

**This is a process measure that is primarily based on two clinical practice guidelines that demonstrate the benefits of systemic corticosteroids and short-acting bronchodilators in the treatment of COPD.

**The evidence applies directly the outcome being measured. Administration of systemic steroids and bronchodilators reduce symptoms and shorten recovery time.

**Event-based process measure, supported by high quality evidence

1b. Performance Gap

Comments:

**There a gap in care that warrants a national performance measure.

**Performance data was provided, utilizing the HEDIS set; developer noted that COPD has high incidence, but did not provide cost impact estimate. Some disparities were noted based on age and race; however, quality of care differences were not noted after adjustments.

**Performance Gap information was provided. Due to the evidenced based data on treatment and treatment guidelines, less than optimal performance results in shorter hospital stays, fewer treatment failures, and improved lung function.

The data on meaningful difference in performance was demonstrated by calculating the inter-quartile range (IQR) for each indicator. This identified a statistically significant 7-16% gap in performance between the 25th and 75th percentile performing plans across the different product lines.

**Remarkable performance gap exists.

1c. High Priority (previously referred to as High Impact)

Comments:

**N/A

**Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

The developer notes the following:

- This measure is a process measure.
- The measure specifies two separate numerators (for systemic corticosteroids and bronchodilators): *Number of patients dispensed a prescription for medication management (on or 14 days after the Episode Date for systemic corticosteroids and on or 30 days after the Episode start Date for bronchodilators) within the measurement year*.
- The denominator specifies two criteria for patients, at least one of which must be met in the measurement year, and the number of events that each patient met based on this criteria: *All patients age 40 years or older as of January 1 of the measurement year with a COPD exacerbation as indicated by an acute inpatient discharge or ED encounter with a principal diagnosis of COPD*. The denominator for this measure is based on acute inpatient discharges and ED visits, not patients.
- ICD-9 and ICD-10 codes have been included in specification details
- The calculation algorithm is provided at S.18 and appears straightforward.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

The developer noted the following:

- The developer conducted beta-binomial at the measure score level. Per the developer, this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that al variation is caused by a real difference in performance (across accountable entities).
- Data were from health plans (241 commercial, 157 Medicaid) that submitted HEDIS data for 2012 and 2015 (2014 measurement year=two years' of health plan data—January 1, 2013 through December 31, 2014); this submission presents the 2015 analysis.
- The beta-binomial method also was used for #0549. Reliability statistics for this measure were calculated using HEDIS health plan data for 2010 and are provided in the following section.

Results of measure score reliability testing

- Per the developer, a minimum reliability score of 0.7 generally is used to indicate sufficient signal strength to discriminate performance between accountable entities.
- Per the developer, the 10-90th percentile distribution of health plan level-reliability on the rates in this measure show the vast majority of health plans exceeded 0.7, and the majority of plans exceeded 0.8. Per the developer, strong reliability is demonstrated since majority of variance is due to signal and not to noise.
- The developer did not present a similar granularity of information for #0549, but did report overall reliability statistics, as noted below. Reliability statistics for #2856 vs. #0549 were similar for Medicaid plans. For commercial plans, reliability statistics were poorer for #2856 (current submission) as compared that for #0549.

The overall beta-binomial statistics for each indicator for commercial and Medicaid plans (and for #0549 Medicare) follow:

- Bronchodilator Indicator: 2015: Commercial = 0.62; Medicaid = 0.94; 2010: Commercial = 0.74; Medicaid = 0.92; Medicare = 0.91
- Systemic Corticosteroid Indicator: 2015: Commercial = 0.56; Medicaid = 0.94; 2010: Commercial = 0.73; Medicaid = 0.93; Medicare = 0.94

Guidance from the Reliability Algorithm: 1→2→3→4→5→6 (highest eligible rating is HIGH)

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Are Committee members concerned about, and therefore wish to discuss with the developer, the decrease in the reliability statistics for commercial plans in the current submission when compared to #0549?*

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**

Specifications not completely consistent with evidence

- Citations specific to the 30-day requirement for dispensing a bronchodilator (or evidence of an active prescription) did not appear to be noted.
- For #0549: The previous Committee expressed concerns about the validity of capturing the numerator, particularly as it relates to EDs—e.g., patients can be told to continue using their existing meds without a new prescription. It also was noted that although the specifications indicate medications dispensed in the ED are provided for and are counted if on claim, there was no way to capture samples being dispensed.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- The developer conducted face validity and empirical validity testing at both the critical data element and performance score levels.
- Face Validity: Measure was tested for face validity using input from 3 expert panels.
- Critical Data Element Validity Testing:
  - Initial empirical testing in 2004 was conducted by comparing the presence of administrative claims codes for patients who had a COPD exacerbation managed in the emergency department or hospital and were discharged home (required to calculate the denominator) to documentation in the medical record, which is considered to be the "gold standard". The plans also looked at administrative claims codes for patients who had a systemic corticosteroid prescription or bronchodilator prescription (required to calculate the numerator) and searched for documentation in the medical record.

- In its submission for #0549, the developer provided supplemental information on data element-level validity to determine the ability to capture COPD exacerbations in administrative claims data. A random sample of 200 charts from 2 different medical centers was reviewed. Some of this material is provided in the current submission.

- Construct Validity Testing (new since #0549): The developer examined whether this measure was correlated with other similar measures of respiratory care. Pearson Correlation Coefficients (PCC) were calculated for data from 241 commercial health plans, 357 Medicare health plans, and 157 Medicaid health plans that submitted data on this measure to HEDIS in 2015, using data from measurement year 2014. The Pearson Correlation test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Validity testing results:

- Face Validity: The developer reports the expert panels agreed that the performance on the indicators were an accurate representation of quality performance and distinguished performance among health plans.
- Critical Data Element Validity (4 plans):
  - In 4 plans, there was 64.2% data consistency for steroid use between administrative and medical record data.
  - There was 66.6% consistency for bronchodilator use between administrative and medical record data
  - Validation of a COPD exacerbation in the medical record was 49%, with a range of 36% to 71%.
  - As part of the request for consideration for #0549, NQF staff calculated the sensitivity, specificity, PPV, and NPV. (As just noted, the current submission presents the data on agreement between the administrative and medical record data, but not these values.) From the data provided by the developer:
    - Sensitivity = 58.2%
    - Specificity = 77.5%
    - Positive Predictive Value = 85.3%
    - Negative Predictive Value = 45.2%

	# exacerbations	Confirm MR and admin	Admin only	MR only	Neither MR nor Admin
Total	159	40.3% (64)	6.9% (11)	28.9% (46)	23.8% (38)

	+Med record	- Med record	Total
+ Admin	64 (True positive)	11 (False positive)	75
- Admin	46 (False negative)	38 (True negative)	84
Total	110	49	159

  - Construct Validity Testing (241 commercial, 357 Medicare, and 157 Medicaid health plan; measurement year 2014):
    - Correlation of a HEDIS spirometry measure to each numerator element and the numerators to each other were analyzed.

    Per the developer:
    - PCCs ranged from 0.1-0.9.
    - The results indicated that the COPD measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized.

Questions for the Committee:

- *Does the Committee have concerns about the validity of the numerator?*

- *Is the sensitivity and specificity of the critical data elements using administrative claims data sufficient for a performance measure used for accountability?*
- *Given the results for the body of validity testing, do the results demonstrate sufficient validity so that conclusions about quality can be made from a computed performance score?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- No exclusion analysis testing conducted
- The developer noted elsewhere, however:
  1. Exclude episode dates when the patient was transferred directly to an acute or non-acute inpatient care setting for any diagnosis. Organizations may identify "transfers" using their own methods and then confirm the acute or non-acute inpatient care setting using codes in the Inpatient Stay Value Set.
  2. Exclude episode dates when the patient was readmitted to an acute or non-acute inpatient care setting for any diagnosis within 14 days after the episode date. To identify readmissions to an acute or non-acute inpatient care setting:
     A. Identify all acute and non-acute inpatient stays (Inpatient Stay Value Set)
     B. Identify the admission date for the stay
  3. Exclude episode dates when the patient had an ED visit (ED value set) for any diagnosis within 14 days after the episode date.

Questions for the Committee:

- *Are the exclusions appropriate?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer notes:

- Meaningful difference in performance was demonstrated by calculating the inter-quartile range (IQR) for each indicator
- Results from 2014 HEDIS data identified a 7-16% gap in performance between the 25th and 75th percentile performing plans across the different product lines and indicators; for all product lines and indicators, the difference between the 25th and 75th percentile performance rates is statistically significant.
- To provide context, the developer estimated that, based on average health plan eligible population, 79 additional members per Medicaid plan would have been discharged on a systemic corticosteroid and 55 additional members would have been discharged on a bronchodilator if plans in the 25th percentile performed as well as plans in the 75th percentile.

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- The developer noted plans collect this measure using all administrative data sources; NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

Guidance from the Validity Algorithm: 1→2→3→6→7→8 (highest eligible rating is HIGH)

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications
Comments:
**Measure 0549 did not meet the threshold of validity due to several concerns with the numerator of this measure. Upon this initial review, I do not share the same concern with the previous committee. The specifications, validity, and reliability all appear appropriate.
**Brief noted that specifications are someone consistent with evidence due to non-reference of 30-day requirement for dispensing a bronchodilator.
**As the prior committee noted, there is no evidence based data to support the 30 day dispensing requirements for bronchodilators. There was a reference to the use of systemic corticosteroids use within 2 weeks of COPD exacerbation. Also as the last committee noted, there may be data collection issues with patients' dispensed medications in the ED.
**There are a number of concerns including capturing dispensed samples, lack of Rx as patient already has medication, and primary non-compliance (i.e., measure relies on dispensed Rx). Adjustment for SDS also should be considered for this measure.

2a2. Reliability Testing
Comments:
**see above
**Validity was tested against both measure score and data element levels with face and empirical validity testing. 200 charts and 700 plans were examined for critical and construct testing, respectively, lending to adequate scope for generalization. Developer indicates that expert panels agreed on performance on indicators representing quality and differentiation between plans.
**Validity scores for Construct validity were above 0.2. There was a lower Sensitivity and Specificity between administrative data and chart data than I would expect. This may be due to not have sufficient records from the hospitals where the majority of treatment occurred. That being said, based upon testing I believe from this data conclusions about quality can be made.
**Exclusions are appropriate. There are no contraindications of which I am aware that would interfere with Rx for corticosteroid or bronchodilators.

2b2. Validity Testing
Comments:
**The developer noted plans collect this measure using all administrative data sources; NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.
**Developer did not perform risk adjustment or exclusion analysis, noting that several exclusions were made, but may vary due to individual methods utilizes by organizations to denote "transfers". Additional exclusions included readmissions and ED visits within 14 days. Developer notes a significant difference in performance (7-16% gap). No comparability or missing data was noted.
**No bias due to missing data

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments:
**see above
**The number of episodes was not noted in the brief, but developer did note use of 398 total health plans. Reliability scores fell mostly above 0.7, which indicates variance due to signal and therefore high reliability.
**Yes there was adequate scope and method. However reliability scores (Beta Binomial) were noted to be under 0.7 for commercial plans (especially for the lower percentile scoring plans) but above 0.7 for Medicare and Medicaid.
**Low sensitivity is a concern for this measure

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer noted:

- All data elements are defined fields in electronic claims and are generated or collected by healthcare personnel during the provision of care. The data are coded by someone other than the individual obtaining the original information.

- NCQA conducts an independent audit of all HEDIS collection and reporting processes to verify that HEDIS specifications are met.

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments:

**Developer noted that all defined fields are available in electronic form, but are not coded by the person obtaining the original data--potential for miscoding?
**Data collection may not be accurate for prescriptions if patient get prescriptions filled outside of the plan and therefore not captured in claims data. For example, prescription filled at the Veterans Administration or the patient does not have a drug prescription benefit with the health plan.
**Measure based on administrative claims data. Low sensitivity merits explanation.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

The developer noted the measure is:

- Publicly reported nationally and by geographic regions in NCQA's *State of Health Care* annual report.
- Reported in *Consumer Reports* and on the NCQA website and used to calculate health plan ratings.
- Used in scoring for accreditation of Medicare Advantage Health Plans
- Used in NCQA's *Quality Compass*.
- Used in the CMS Medicare Advantage Star Rating program

Improvement results

The developer notes:

- From 2012-2014, the average rate on the systemic corticosteroid indicator showed improvement across Commercial PPO and Medicare HMO and PPO plans (see section 1b.2). There also was improvement in plans at the 90th percentile for commercial, Medicare and Medicaid plans on the systemic corticosteroid indicator; these data are nationally representative.

Unexpected findings (positive or negative) during implementation

- The developer states there were no identified unintended consequences for this measure during testing or since implementation.

Potential harms

- The developer states there were no identified unintended consequences for this measure during testing or since implementation.

Feedback : No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Question for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments:

**Currently publicly reported and used for accountability purposes.
**Measure is currently being publicly reported as part of NCQA's State of Health Care annual report, Consumer Reorts, NCQA website, calculations of health plan ratings, accreditation of Medicare Advantage health plans, Quality Compass, and in the CMS Medicare Advantage Star Rating program. Developer states that use of measure during that time has shown improvement between 2012 and 2014 on corticosteroid indicators, potential exists to measure to close gap between private and public payers as well. No unintended consequences are noted, but attention should be given to ensure disparate populations have access to providers and adequate treatment as appropriate.
**This measure can be used to provide feedback to providers on adherence to evidenced based treatment guidelines for COPD exacerbations.
**Yes.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0102: COPD: inhaled bronchodilator therapy

Harmonization

NCQA provides the following response as to why NQF #2856 is not harmonized:

- NQF #0102 is a physician-level measure and the focus of this proposed measure is different. NQF #2856 focuses exclusively on patients who were hospitalized or had an ED visit for a COPD exacerbation and received timely recommended treatment (systemic corticosteroids and bronchodilators), while NQF # 0102 focuses on managing COPD and allows receipt of a bronchodilator at least once during the measurement year.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Pharmacotherapy Management of COPD Exacerbation

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Pharmacotherapy Management of COPD Exacerbation

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Patient has inpatient or ED visit for a COPD exacerbation >>> Patient receives systemic corticosteroid and bronchodilator to treat symptoms >>> Patient has improved outcomes including improved lung function and arterial hypoxemia, reduced risk of early relapse or treatment failure, and shorter recovery time and length of hospital stay

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines:

Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Bethesda (MD): Global Initiative for Chronic Obstructive Lung Disease, 2015.

http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html

Institute for Clinical Systems Improvement (ICSI) Guidelines:

Anderson, B., K. Conner, C. Dunn, et al. 2013. Institute for Clinical Systems Improvement. Diagnosis and Management of Chronic Obstructive Pulmonary Disease (COPD). https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_respiratory_guidelines/copd/

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

GOLD GUIDELINES, MANAGEMENT OF EXACERBATIONS, page 42

Pharmacologic Treatment

- *Short-acting Bronchodilators:* Although there are no controlled trials, short-acting inhaled beta$_2$-agonists with or without short-acting anticholinergics are usually the preferred bronchodilators for treatment of an exacerbation (Evidence C).
- *Corticosteroids:* Data from studies in secondary health care indicate that systemic corticosteroids in COPD exacerbations shorten recovery time, improve lung function (FEV_1) and arterial hypoxemia (PaO_2) (Evidence A), and reduce the risk of early relapse, treatment failure, and length of hospital stay. A dose of 40 mg prednisone per day for 5 days is recommended (Evidence B).

ICSI Guidelines, page 26-27:

- **Bronchodilators:** Albuterol and levalbuterol are the preferred bronchodilators in the setting of an acute exacerbation of COPD because of their rapid onset of action. If clinical improvement does not occur promptly, ipratropium may be added to produce additive bronchodilation and allow the use of lower doses of albuterol or levabuterol, thus diminishing dose-dependent toxicity. Administration of either agent by metered-dose inhaler and spacer or by nebulization is acceptable *(Turner, 1997 [Meta-analysis]; Moayyedi, 1995 [High Quality Evidence]; Patrick, 1990 [High Quality Evidence]).*
- **Systemic Steroids:** Studies have demonstrated the benefits of systemic glucocorticosteroids in the management of COPD exacerbations. Doses of oral prednisone at 30 to 40 mg a day for 7 to 14 days have been shown to reduce symptoms and reduce the likelihood of hospitalization. Treatment beyond two weeks does not provide any additional benefit, but does increase the likelihood of significant side effects such as hyperglycemia. There is no need to discontinue inhaled steroids while the patient is taking oral prednisone (*Aaron, 2003 [High Quality Evidence]; McEvoy, 2000 [Low Quality Evidence]; Davies, 1999 [High Quality Evidence]; Niewoehner, 1999 [High Quality Evidence]; Thompson, 1996 [High Quality Evidence]).*

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Levels of Evidence:

Evidence A. Randomized controlled trials. Rich body of data. Evidence is from endpoints of well-designed randomized controlled trials that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.

Evidence B. Randomized controlled trials. Limited data. Evidence is from endpoints of intervention studies that include only a limited number of patients, posthoc or subgroup analysis of randomized controlled trials, or meta analysis of randomized controlled trials. In general, Category B pertains when few randomized trials exist, they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.

Evidence C. Nonrandomized trials. Observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

ICSI Guidelines:

- High Quality Evidence = Further research is very unlikely to change our confidence in the estimate of effect.
- Low Quality Evidence = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain.
- Meta-analysis

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Levels of Evidence:

Evidence D. Panel consensus. Judgment. This category is used only in cases where the provision of some guidance was deemed valuable but the clinical literature addressing the subject was deemed insufficient to justify placement in one of the other categories. The Panel Consensus is based on clinical experience or knowledge that does not meet the above listed criteria.

ICSI Guidelines:

- Moderate Quality Evidence = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- Guideline
- Systematic review
- Decision-analysis
- Cost-effectiveness analysis

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Bethesda (MD): Global Initiative for Chronic Obstructive Lung Disease, 2015.

http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html

Institute for Clinical Systems Improvement (ICSI):

Anderson, B., K. Conner, C. Dunn, et al. 2013. Institute for Clinical Systems Improvement. Diagnosis and Management of Chronic Obstructive Pulmonary Disease (COPD). https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_respiratory_guidelines/copd/

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

This measure assesses whether patients who were hospitalized or had an emergency department (ED) visit for a COPD exacerbation were dispensed appropriate medication to treat their symptoms. This measure is based on evidence that systemic corticosteroids and bronchodilators following a COPD exacerbation have been shown to improve symptoms, reduce the risk of early relapse, and shorten recovery time and length of hospital stay.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Levels of Evidence:

Evidence A. Randomized controlled trials. Rich body of data. Evidence is from endpoints of well-designed randomized controlled trials that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.

Evidence B. Randomized controlled trials. Limited data. Evidence is from endpoints of intervention studies that include only a limited number of patients, posthoc or subgroup analysis of randomized controlled trials, or meta analysis of randomized controlled trials. In general, Category B pertains when few randomized trials exist, they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.

Evidence C. Nonrandomized trials. Observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

ICSI Guidelines:

- High Quality Evidence = Further research is very unlikely to change our confidence in the estimate of effect.
- Low Quality Evidence = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain.
- Meta-analysis

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Levels of Evidence:

Evidence D. Panel consensus. Judgment. This category is used only in cases where the provision of some guidance was deemed valuable but the clinical literature addressing the subject was deemed insufficient to justify placement in one of the other categories. The Panel Consensus is based on clinical experience or knowledge that does not meet the above listed criteria.

ICSI Guidelines:

- Moderate Quality Evidence = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- Guideline
- Systematic review
- Decision-analysis
- Cost-effectiveness analysis

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range:

GOLD Guidelines: 1965-2014

ICSI Guidelines: 1968 – 2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The guideline developers did not provide a breakdown of the specific number of randomized control trials (RCTs) and given the number of studies included in the systematic reviews we were not able to delineate all RCTs for each recommendation. The GOLD guidelines referenced a total of 613 studies to update the previous set of guidelines from 2013. The recommendation related to systemic corticosteroid and short-acting bronchodilators to treat a COPD exacerbation was based on 25 studies, including randomized control trials (RCTs), meta-analyses, systematic reviews of randomized controlled trials (RCTs), and observational studies. The ICSI guideline recommendations related to short-acting bronchodilators following an exacerbation referenced one meta-analysis of 12 RCTs. The ICSI guideline recommendations related to systemic corticosteroids following an exacerbation referenced four randomized control trials and one observational study.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall, the quality of the evidence that systemic corticosteroids improve outcomes following a COPD exacerbation is high. The evidence is based on the results of multiple randomized control trials conducted in hundreds of patients with COPD. These studies have consistently found that patients who received systemic corticosteroids within two weeks following an exacerbation had improved lung function, fewer treatment failures, increased time to relapse, and shorter hospital stays than patients who did not receive systemic

corticosteroids. The quality of evidence that short-acting bronchodilators improve outcomes following COPD exacerbation is moderate as the RCTs were conducted in fewer patients; however, the findings from the meta-analysis of 12 RCTs found that short-acting bronchodilators caused statistically significant increases in pulmonary function tests and are important for rapidly improving shortness of breath and wheezing.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Studies have consistently found that the primary benefit of short-acting bronchodilators for COPD patients is increasing FEV_1 and thus are recommended as as-needed therapy including following a COPD exacerbation. Randomized control trials consistently found that patients who received systemic corticosteroids within two weeks after experiencing a COPD exacerbation had statistically significant better outcomes compared to patients who did not. The following are examples of statistically significant study findings showing the magnitude of the outcomes:

- The percentage-predicted FEV1 after bronchodilation rose more frequently in corticosteroid-treated patients (from 28% to 42%) compared to - patients not treated with corticosteroid (from 26% to 32%).
- Systemic glucocorticoids were associated with a shorter initial hospital stay (8.5 days, vs. 9.7 days for placebo).
- The overall rate of relapse at 30 days was lower in the corticosteroid group than in the placebo group (27 percent vs. 43 percent).

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

One risk associated with the use of bronchodilators is toxicity, however the benefit of use during an acute episode outweighs the risk. The risk of toxicity is mainly attributed in long term medication use in stable COPD. Patients who receive consistent systematic corticosteroid therapy are at higher risk for hyperglycemia or osteoporosis. However, there is consensus that short-term use of systemic corticosteroids to treat a COPD exacerbation outweigh the potential harms.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PCE_Evidence.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This measure assesses whether patients who had a hospitalization or an emergency department (ED) visit for a COPD exacerbation were provided appropriate medication (systemic corticosteroids and bronchodilators) to treat symptoms and prevent future exacerbations. The improvement in quality envisioned by the use of this measure is to increase the use of systemic corticosteroids and bronchodilators following a COPD exacerbation in order to shorten patients' recovery time, improve lung function and reduce the risk of early relapse, treatment failure, and length of hospital stay.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year and product line (i.e. commercial, Medicaid, and Medicare).

Systemic corticosteroids Commercial HMO Rate

YEAR	MEAN	ST DEV	MIN	10TH	25TH	50TH	75TH	90TH	MAX	Interquartile Range
2012	74%	8%	53%	64%	69%	74%	78%	81%	91%	9%
2013	76%	7%	58%	67%	71%	76%	80%	83%	91%	8%
2014	75%	7%	50%	68%	72%	76%	80%	84%	88%	8%

Systemic corticosteroids Commercial PPO Rate

2012	71%	8%	46%	60%	66%	71%	76%	79%	90%	10%
2013	73%	7%	51%	64%	69%	73%	77%	80%	92%	8%
2014	73%	7%	48%	64%	69%	74%	78%	82%	85%	9%

Systemic corticosteroids Medicare HMO Rate

YEAR	MEAN	ST DEV	MIN	10TH	25TH	50TH	75TH	90TH	MAX	Interquartile Range
2012	69%	10%	28%	58%	66%	71%	75%	79%	94%	8%
2013	71%	9%	33%	60%	67%	73%	76%	79%	87%	9%
2014	72%	9%	29%	61%	69%	74%	77%	81%	89%	8%

Systemic corticosteroids Medicare PPO Rate

2012	70%	8%	33%	62%	66%	71%	76%	77%	84%	9%
2013	72%	8%	40%	63%	69%	73%	76%	79%	90%	8%
2014	73%	10%	32%	65%	70%	74%	78%	82%	90%	8%

Systemic corticosteroids Medicaid Rate (HMO and PPO Combined)

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 65% | 12% | 28% | 49% | 61% | 67% | 73% | 77% | 86% | 12%
2013 | 66% | 13% | 8% | 48% | 61% | 69% | 75% | 78% | 90% | 14%
2014 | 65% | 14% | 8% | 48% | 59% | 69% | 75% | 78% | 95% | 16%

Bronchodilator Commercial HMO Rate
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 81% | 6% | 56% | 73% | 77% | 81% | 86% | 88% | 95% | 9%
2013 | 80% | 7% | 56% | 71% | 76% | 81% | 85% | 88% | 95% | 9%
2014 | 81% | 7% | 54% | 73% | 77% | 81% | 85% | 89% | 94% | 8%

Bronchodilator Commercial PPO Rate
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 78% | 7% | 49% | 69% | 74% | 78% | 82% | 86% | 100% | 8%
2013 | 79% | 6% | 59% | 71% | 74% | 78% | 83% | 86% | 91% | 9%
2014 | 78% | 7% | 56% | 68% | 74% | 78% | 82% | 85% | 93% | 8%

Bronchodilator Medicare HMO Rate
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 80% | 8% | 39% | 71% | 77% | 81% | 85% | 90% | 96% | 8%
2013 | 81% | 7% | 47% | 71% | 77% | 81% | 86% | 89% | 97% | 8%
2014 | 81% | 8% | 48% | 72% | 78% | 82% | 86% | 90% | 98% | 8%

Bronchodilator Medicare PPO Rate
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 77% | 7% | 47% | 69% | 73% | 77% | 80% | 84% | 93% | 7%
2013 | 77% | 7% | 48% | 69% | 74% | 77% | 81% | 84% | 93% | 8%
2014 | 78% | 8% | 44% | 68% | 75% | 78% | 82% | 87% | 94% | 7%

Bronchodilator Medicaid Rate (HMO and PPO Combined)
YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range
2012 | 81% | 9% | 45% | 72% | 78% | 83% | 87% | 90% | 94% | 9%
2013 | 81% | 12% | 9% | 69% | 78% | 84% | 88% | 90% | 100% | 10%
2014 | 79% | 13% | 23% | 64% | 76% | 83% | 87% | 89% | 93% | 11%

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Commercial HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 113 | 131
2013 | 123 | 125
2014 | 115 | 119

Commercial PPO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 125 | 190
2013 | 124 | 198
2014 | 126 | 184

Medicare HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 272 | 339
2013 | 267 | 375
2014 | 263 | 386

Medicare PPO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 117 | 296
2013 | 106 | 390
2014 | 94 | 463

Medicaid HMO
YEAR | N Plans | Mean Denominator Size per plan
2012 | 131 | 405
2013 | 142 | 434
2014 | 157 | 497

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
HEDIS data are stratified by type of insurance (e.g. Commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.
Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts. Health Affairs 20(10): 1984-1991.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Healthy People 2020 data show that, since 2007, white adults over age 45 have had a higher mortality rate than other races and ethnicities while black adults over age 45 have experienced higher hospitalization and higher emergency department (ED) utilization (U.S. Department of Health and Human Services 2013). One study found that while there was variation in utilization, there was little variation in short-term outcomes following acute exacerbation (Tsai et al. 2009). Predictors of increased ED visits were male sex, African American race, Medicaid insurance, low socioeconomic status, and certain comorbidities, and patients with high numbers of ED visits had higher frequency of a repeat admission within 30 days (Tsai et al. 2009). When in the ED, African American and Hispanic patients were less likely to receive certain services, such as chest radiography, but were more likely to receive systemic corticosteroids both in the ED and at discharge; however, other disease and symptom management in the ED was not found to be significantly different between races (Tsai et al. 2009). After adjusting for patient and ED characteristics, quality of care was not found to differ significantly between different races and ethnicities (Tsai et al. 2009).

Tsai, C.L., and C.A. Camargo, Jr. 2009. Racial and ethnic differences in emergency care for acute exacerbation of chronic obstructive pulmonary disease. Academic Emergency Medicine. 16(2):108-115. doi: 10.1111/j.1559-2712.2008.00319.x.

U.S. Department of Health and Human Services. 2013. Respiratory Diseases. http://www.healthypeople.gov/2020/topicsobjectives2020/ (Accessed July 23, 2014).

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, High resource use, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

The National Heart, Lung, and Blood Institute (NHLBI) states that over 15 million adults have been diagnosed with chronic obstructive pulmonary disease (COPD), and that the actual number of those with the disease may be higher (NHLBI 2013). While other major causes of death have been decreasing, COPD mortality has risen, making it the third leading cause of death in the U.S. (Hoyert and Xu 2012; NHLBI 2013). Without intervention, deaths from COPD are projected to increase by more than 30 percent in the next ten years, and COPD is projected to be the third leading cause of death worldwide (WHO 2014). In 2010, COPD was the main cause for 10.3 million physician visits, 1.5 million emergency department visits and 699,000 discharges (Ford et al, 2013).

Exacerbations may be the most significant drivers of negative impacts on a COPD patient (GOLD 2013). Patients experiencing exacerbations are at higher risk for repeat exacerbations, more rapid decline in lung function, and reduced exercise capacity (Donaldson et al. 2002) and these effects are more pronounced for patients with severe COPD (Spencer et al. 2004). In addition to physical effects, COPD exacerbations result in reduced quality of life and ability to conduct activities of daily living independently (Spencer, 2004; Miravitlles et al. 2004). However, proper therapy following an exacerbation, including pharmacotherapy, can slow disease progression and reduce the risk of future exacerbations (GOLD 2015).

Approximately 641,000 hospital inpatient discharges and 294,000 emergency department visits in 2010 were due to COPD (CDC 2014), and the projected total cost of COPD was $49.9 billion, including $29.5 billion for direct health care expenditures (NHLBI 2013). While not all are reported, exacerbations typically result in hospitalization, and the cost of an emergency department visit, simple admissions, and complex admissions can be significant at an average of $647, $7242, and $20,757 each, respectively, in 2008 (Dalal et al. 2011). Exacerbations are the costliest aspect of COPD, and cost of care tends to increase with disease severity (Pasquale et al. 2012). Because COPD exacerbation increases the likelihood of future exacerbations, and hospital readmissions following an emergency department or admission are more costly, disease control and management are essential to reducing both utilization and cost.

1c.4. Citations for data demonstrating high priority provided in 1a.3

Centers for Disease Control and Prevention. 2014. FastStats: Chronic Obstructive Pulmonary Disease (COPD) Includes: Chronic Bronchitis and Emphysema. http://www.cdc.gov/nchs/fastats/copd.htm (Accessed July 23, 2014).

Dalal, A.A., M. Shah, A.O. D'Souza, and P. Rane. 2011. "Costs of COPD exacerbations in the emergency department and inpatient setting." Respiratory Medicine. 105(3):454-60.

Donaldson, G.C., T.A.R. Seemungal, A. Bhowmik, and J.A. Wedzicha. 2002. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. 57:847-852.

Ford E, Croft J, Mannino D, Wheaton A, Zhang X, and Giles W. COPD Surveillance—United States, 1999-2011. Chest 2013; 144(1):284–305.

Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2015. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html (Accessed November 16, 2015).

Hoyert, D., and J. Xu. 2012. Deaths: Preliminary Data for 2011. National Vital Statistics Reports. 61(6):1-52. http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_06.pdf (Accessed July 23, 2014)

Miravitlles, M., M. Ferrer, A. Pont, et al. 2004. Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary

disease: a 2 year follow up study. Thorax. 59:387-395. doi: 10.1136/thx.2003.008730.

National Heart, Lung, and Blood Institute. 2013. Morbidity & Mortality: 2013 Chart Book on Cardiovascular, Lung and Blood Diseases. Available at: http://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook.pdf. (Accessed November 18, 2015).

Pasquale, M.K., S.X. Sun, F. Song, H.J. Hartnett, and S.A. Stemkowski. 2012. Impact of exacerbations on health care cost and resource utilization in chronic obstructive pulmonary disease patients with chronic bronchitis from a predominantly Medicare population. International Journal of COPD. 7:757-64. doi: 10.2147/COPD.S36997

Spencer, S., P.M.A. Calverley, P.S. Burge, and P.W. Jones. 2004. Impact of preventing exacerbations on deterioration of health status in COPD. European Respiratory Journal. 23:698-702.

World Health Organization (WHO). 2014. Chronic Respiratory Diseases: Chronic Obstructive Pulmonary Disease. http://www.who.int/respiratory/copd/en/ (Accessed July 23, 2014).

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and 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. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Pulmonary/Critical Care, Pulmonary/Critical Care : Chronic Obstructive Pulmonary Disease (COPD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
N/A

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** XXXX_PCE_Value_Sets.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population,*

i.e., cases from the target population with the target process, condition, event, or outcome)
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Numerator 1 (Systemic Corticosteroids): The number of patients dispensed a prescription for systemic corticosteroid on or 14 days after the Episode Date*. Count systemic corticosteroids that are active on the relevant date.

Numerator 2 (Bronchodilator): The number of patients dispensed a prescription for a bronchodilator on or 30 days after the Episode Date*. Count bronchodilators that are active on the relevant date.

*The Episode Date is the date of service for any acute inpatient discharge or ED claim/encounter during the 11-month intake period with a principal diagnosis of COPD.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: a 12-month period that begins on January 1 and ends on December 30 of the measurement year.
Denominator: an 11-month period that begins on January 1 and ends on November 30 of the measurement year.

S.6. **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, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Follow the steps below to identify numerator compliance.

Numerator 1 (Systemic Corticosteroid): Identify the number of patients dispensed a prescription for systemic corticosteroid (refer to PCE-C: Systemic Corticosteroids) on or 14 days after the Episode Date.
-The Episode Date is the date of service for any acute inpatient discharge or ED claim/encounter during the 11-month intake period with a principal diagnosis of COPD.
-Count systemic corticosteroids that are active on the relevant date. An active prescription is considered active if the "days supply" indicated on the date the patient filled the prescription is the number of days or more between that date and the relevant date. For an acute inpatient encounter, the relevant date is the date of admission. For an ED claim/encounter, the relevant date is the date of service.

Numerator 2 (Bronchodilator): Identify the number of patients dispensed a prescription for bronchodilator (refer to PCE-D: Bronchodilators) on or 30 days after the Episode Date.
-The Episode Date is the date of service for any acute inpatient discharge or ED claim/encounter during the 11-month intake period with a principal diagnosis of COPD.
-Count bronchodilators that are active on the relevant date. An active prescription is considered active if the "days supply" indicated on the date the patient filled the prescription is the number of days or more between that date and the relevant date. For an acute inpatient encounter, the relevant date is the date of admission. For an ED claim/encounter, the relevant date is the date of service.

PCE-C: Systemic Corticosteroids:
Glucocorticoids: betamethasone, dexamethasone, hydrocortisone, methylprednisolone, prednisolone, prednisone, triamcinolone

PCE-D: Bronchodilators:
Anticholinergic agents: albuterol-ipratropium, aclidinium-bromide, ipratropium, tiotropium, Umeclidinium
Beta 2-agonists: albuterol, arformoterol, budesonide-formoterol, fluticasone-salmeterol, fluticasone-vilanterol, formoterol, Indacaterol, levalbuterol, Mometasone-formoterol, metaproterenol, Olodaterol hydrochloride, pirbuterol, salmeterol, Umeclidinium-vilanterol
Methlyxanthines: aminophylline, dyphylline, dyphylline-guaifenesin, guaifenesin-theophylline, theophylline

See corresponding Excel file for value sets referenced above.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients age 40 years or older as of January 1 of the measurement year with a COPD exacerbation as indicated by an acute inpatient discharge or ED encounter with a principal diagnosis of COPD.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The eligible population for this measure is based on acute inpatient discharges and ED visits, not on patients. It is possible for the denominator to include multiple events for the same individual. The eligible population for the denominator is defined by following the series of steps below:

Step 1: Identify all patients who had either of the following during the Intake Period (an 11-month period that begins on January 1 of the measurement year and ends on November 30 of the measurement year):
1) An ED visit (ED Value Set) with a primary diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include ED visits that result in an inpatient admission.
2) An acute inpatient discharge with a primary diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). To identify acute inpatient discharges:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)
c. Identify the discharge date for the stay

Step 2: Identify all COPD Episode Dates (the date of service for any acute inpatient discharge or ED claim/encounter during the intake period with a principal diagnosis of COPD). For each patient in Step 1, identify all acute inpatient discharges and ED Visits.

See corresponding Excel file for value sets referenced above.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
1) Exclude episode dates when the patient was transferred directly to an acute or nonacute inpatient care setting for any diagnosis.
2) Exclude episode dates when the patient was readmitted to an acute or nonacute inpatient care setting for any diagnosis within 14 days after the episode date.
3) Exclude episode dates when the patient had an ED visit for any diagnosis within 14 days after the Episode date.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1) Exclude episode dates when the patient was transferred directly to an acute or nonacute inpatient care setting for any diagnosis. Organizations may identify “transfers” using their own methods and then confirm the acute or nonacute inpatient care setting using codes in the Inpatient Stay Value Set.

2) Exclude episode dates when the patient was readmitted to an acute or nonacute inpatient care setting for any diagnosis within 14 days after the episode date. To identify readmissions to an acute or nonacute inpatient care setting:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
b. Identify the admission date for the stay

3) Exclude episode dates when the patient had an ED visit (ED value set) for any diagnosis within 14 days after the episode date.

See corresponding Excel file for value sets referenced above.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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 = Higher score

S.18. **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.)*
Refer to items S.6 (Numerator details), S.9 (Denominator details), S.11 (Denominator exclusions details) and S.2b (Data Dictionary) for tables.

The denominator for this measure is based on acute inpatient discharges and ED visits, not patients. The measure calculation is detailed in the steps listed below:

Step 1: identify the eligible population.

A. Identify all patients who had either of the following during the Intake Period (an 11-month period that begins on January 1 of the measurement year and ends on November 30 of the measurement year):
1) An ED visit (ED Value Set) with a primary diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include ED visits that result in an inpatient admission.
2) An acute inpatient discharge with a primary diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). To identify acute inpatient discharges:
a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)
c. Identify the discharge date for the stay

B. Identify all COPD Episode Dates (the date of service for any acute inpatient discharge or ED claim/encounter during the intake period with a principal diagnosis of COPD). For each patient in Step 1, identify all acute inpatient discharges and ED Visits.

Step 2: determine denominator exclusions.

A. Exclude episode dates when the patient was transferred directly to an acute or nonacute inpatient care setting for any diagnosis. Organizations may identify “transfers” using their own methods and then confirm the acute or nonacute inpatient care setting using codes in the Inpatient Stay Value Set.
B. Exclude episode dates when the patient was readmitted to an acute or nonacute inpatient care setting for any diagnosis within 14 days after the episode date. To identify readmissions to an acute or nonacute inpatient care setting:
1. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)
2. Identify the admission date for the stay
3. Exclude episode dates when the patient had an ED visit (ED value set) for any diagnosis within 14 days after the episode date.

Step 3: determine the numerator.

Numerator 1 (Systemic Corticosteroid): Identify the number of patients dispensed a prescription for systemic corticosteroid (refer to PCE-C: Systemic Corticosteroids) on or 14 days after the Episode Date.
-The Episode Date is the date of service for any acute inpatient discharge or ED claim/encounter during the 11-month intake period with a principal diagnosis of COPD.
-Count systemic corticosteroids that are active on the relevant date. An active prescription is considered active if the "days supply" indicated on the date the patient filled the prescription is the number of days or more between that date and the relevant date. For an acute inpatient encounter, the relevant date is the date of admission. For an ED claim/encounter, the relevant date is the date of service.

Numerator 2 (Bronchodilator): Identify the number of patients dispensed a prescription for bronchodilator (refer to PCE-D: Bronchodilators) on or 30 days after the Episode Date.
-The Episode Date is the date of service for any acute inpatient discharge or ED claim/encounter during the 11-month intake period with a principal diagnosis of COPD.
-Count bronchodilators that are active on the relevant date. An active prescription is considered active if the "days supply" indicated on the date the patient filled the prescription is the number of days or more between that date and the relevant date. For an acute inpatient encounter, the relevant date is the date of admission. For an ED claim/encounter, the relevant date is the date of service.

Step 4: calculate two rates.
A. Number of patients dispensed a prescription for systemic corticosteroid on or 14 days after the Episode Date/Denominator
B. Number of patients dispensed a prescription for bronchodilator on or 30 days after the Episode Date /Denominator

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PCE_Testing.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: **Pharmacotherapy Management of COPD Exacerbation**
Date of Submission: **12/14/2015**
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed

performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score

as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
N/A

1.3. What are the dates of the data used in testing? Click here to enter date range
Initial testing: During measure development, we conducted a comprehensive field test in 2004 to assess feasibility of data collection and validity of performance data and critical data elements. This field test used data from measurement year 2003, which included health plan data spanning December 1, 2002 through January 31, 2004.

Systematic evaluation of face validity: The measure was tested for face validity throughout measure development from 2004 to 2006.

Measure score reliability and construct validity testing: We assessed measure score reliability and construct validity using data from all health plans that submitted HEDIS data to NCQA for this measure in 2012 and again in 2015. In this form, we provide the 2015 measure score reliability and construct validity testing results, which used data for measurement year 2014. Measurement year 2014 required health plan data from January 1 through December 31, 2014.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Initial testing: To assess feasibility of data collection and validity of performance data and critical data elements, 5 Commercial health plans, 1 Medicaid health plan and 3 Medicare health plans provided individual member-level data to NCQA for analysis. These plans were selected because they had the resources to generate the files, had sufficient sample of members with persistent asthma for analysis, and willingness to provide the data. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity: Throughout the entire measure development process from 2004-2006, the measure was tested for face validity using panels of experts with specific clinical, methodologic and operational expertise. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliation of expert panel:

1) NCQA's Respiratory Measurement Advisory Panel (RMAP) is comprised of 10 experts (8 physicians, 1 pharmacist and 1 researcher) in clinical pulmonary care, including health care providers and policy makers.

2) NCQA's Technical Measurement Advisory Panel is a 12-member panel representing health plans methodologists, clinicians and HEDIS auditors.

3) NCQA's Committee on Performance Measurement (CPM) oversees the HEDIS measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 17 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

4) NCQA's HEDIS Expert Coding Panel reviewed and provided feedback on the vocabularies and definitions found in the values sets used to identify each measure component as well as the more recent mapping of ICD-9 codes to ICD-10 codes.

In 2005, the draft measure was posted for public comment, a 30-day period of review that allowed interested parties to offer feedback to NCQA about the measure. Stakeholders from various types of organizations submitted 67 comments on the measure.

Measure score reliability and construct validity testing: Measure score reliability and construct validity was calculated from the 241 Commercial health plans (comprising 115 HMOs and 126 PPOs), 357 Medicare plans (comprising 263 HMOs and 94 PPOs) and 157 Medicaid health plans that submitted data on this measure to HEDIS in 2015. The plans were geographically diverse and varied in size.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Patient sample for initial measure field testing: We collected data from 5 Commercial health plans, 3 Medicare health plans and 1 Medicaid health plan. Below is a description of the sample. It includes the number of health plans that provided data for the measurement year 2003 and the median denominator for the measure across health plans. Note that the denominator is based on acute inpatient discharges and ED visits for COPD exacerbations, not members.

Product Type	Number of Plans	Median Number of Hospital/ED Visits per Plan
Commercial	5	86
Medicare	3	274
Medicaid	1	690

Measure score reliability and construct validity testing: In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the number of health plans that submitted data for this measure to HEDIS for measurement year 2014 and the median denominator for the measure across health plans. Note that the denominator is based on acute inpatient discharges and ED visits for COPD exacerbations, not members.

Product Type	Number of Plans	Median Number of Hospital/ED Visits per Plan
Commercial HMO	115	79
Commercial PPO	126	87
Medicare HMO	263	187
Medicare PPO	94	142
Medicaid	157	350

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

During measure development, we conducted a comprehensive field test in 2004 to assess feasibility of data collection and validity of performance data and critical data elements using data submitted by 5 Commercial health plans, 1 Medicaid health plan and 3 Medicare health plans. The field test used data from measurement year 2003, which included health plan data spanning December 1, 2002 through January 31, 2004.

Face validity was demonstrated through a systematic assessment of face validity during measure development. Per NQF instructions we have described the composition of the technical expert panel, which assessed face validity in the data sample questions above.

The measure underwent additional analyses to assess measure score reliability (tested using a beta-binomial calculation) and construct validity (tested using Pearson's correlations of similar measures). These analyses included all of the health plans (241 Commercial, 357 Medicare, and 157 Medicaid) that submitted data for this measure to HEDIS for measurement year 2014 (measurement year 2014 required health plan data from January 1 through December 31, 2014).

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Measure performance results are stratified by Commercial, Medicare and Medicaid health plans.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Beta-Binomial Method:

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Beta-Binomial Statistic:

	Commercial			Medicare			Medicaid		
	Median	Overall	10th-90th	Median	Overall	10th-90th	Median	Overall	10th-90th
Bronchodilator Indicator	0.61	0.62	0.41-0.84	0.86	0.82	0.62-0.97	0.97	0.94	0.83-0.99
Systemic Corticosteroid Indicator	0.55	0.56	0.35-0.81	0.85	0.82	0.61-0.97	0.97	0.94	0.85-0.99

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Interpretation of measure score reliability testing:

Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The results of this beta binomial testing suggest that this measure has good reliability. The 10-90th percentile distribution of health plan level reliability on this measure show the vast majority of health plans exceeded the minimally accepted threshold of 0.7, and the majority of plans exceeded 0.8. Strong reliability is demonstrated since the majority of variances is due to signal and not to noise.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

METHOD OF ASSESSING FACE VALIDITY: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

This measure was developed in 2004 to assess whether patients who had a hospitalization or an emergency department visit for a COPD exacerbation were provided appropriate medication (systemic corticosteroids and bronchodilators) to treat symptoms and prevent future exacerbations. NCQA and the Respiratory Measurement Advisory Panel worked together to develop the most appropriate measure for assessing medication management of COPD exacerbations.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

The pharmacotherapy for COPD exacerbation measure was written and field-tested in 2004. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in 2005.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

The pharmacotherapy for COPD exacerbation measure was released for Public Comment in 2005 prior to publication in HEDIS. We received and responded to 67 comments on this measure. The CPM recommended moving this measure to first year data collection by a majority vote.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected,

reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

The pharmacotherapy for COPD exacerbation measure was introduced to HEDIS in 2005. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure to public reporting with a majority vote in 2006.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publically reported and may be used for scoring in accreditation.

The pharmacotherapy for COPD exacerbation measure has been publicly reported in HEDIS since 2006.

STEP 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

The clinical guideline recommendations for pharmacotherapy management following COPD exacerbation have not changed since the measure was developed in 2005; therefore, we have not made any significant changes to the measure since it was last endorsed on January 31, 2012.

Expert Participation

This measure was tested for face validity with input from three expert panels. Guidelines from the Global Initiative for Chronic Obstructive Lung Disease (GOLD) were also a strong authoritative source in applying the evidence for the measure.

We list an overview of each panel here. Please refer to Ad.1 in the submission form for the names and affiliation of experts in each panel.

1) Respiratory Measurement Advisory Panel includes 10 members (eight physicians, one pharmacist and a researcher) with expertise in respiratory care and quality measurement.
2) The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.
3) NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

ICD-10 CONVERSION:
In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA first identified value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM)to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.
3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented to for expert review and feedback.
4. NCQA RMAP clinical review: Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.
5. New value sets containing ICD-10 code recommendations were for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under **Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development**.

METHOD OF TESTING CRITICAL DATA ELEMENT VALIDITY: Validity was tested by comparing the presence of administrative claims codes for patients who had a COPD exacerbation managed in the emergency department or hospital and were discharged home (required to calculate the denominator) to documentation in the medical record, which is considered to be the "gold standard". The plans also looked at administrative claims codes for patients who had a systemic corticosteroid prescription or bronchodilator prescription (required to calculate the numerator) and searched for documentation in the medical record.

METHOD OF TESTING CONSTRUCT VALIDITY: We tested for construct validity by exploring whether this measure was correlated with other similar measures of respiratory care. We hypothesized that organizations that perform well on the measure should perform well on other similar HEDIS measures. To test these correlations we used a Person correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

For this measure, we specifically hypothesized:
1) Performance on the systemic corticosteroid indicator (percent of patients who were dispensed a prescription for a systemic corticosteroid within 14 days after an acute COPD exacerbation) will be positively correlated with the measure assessing whether patients received a spirometry test to confirm COPD diagnosis.
2) Performance on the bronchodilator indicator (percent of patients who were dispensed a prescription for a bronchodilator within 30 days after an acute COPD exacerbation) will be positively correlated with the spirometry measure.
3) Performance on the systemic corticosteroid indicator will be positively correlated with the bronchodilator indicator.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

RESULTS OF FACE VALIDITY ASSESSMENT:
For the initial field test conducted in 2004, we calculated the total measure performance rate, as well as performance rates based on whether patients had an active prescription that counted toward the numerator or received a new prescription. We discussed the validity of the performance results with the expert panels (Respiratory Measurement Advisory Panel and the Committee on Performance Measurement). Among the different product lines, performance rates ranged from 36 to 47 percent for the systemic corticosteroid indicator and 45 to 64 percent for the bronchodilator indicator. Between 9 to 14 percent of the discharged members had an active systemic corticosteroid prescription and 26 to 40 percent of discharged members had an active bronchodilator prescription. Performance rates

on both indicators were slightly lower for men compared to women and for people ages 75+ compared to ages 40-74. The expert panels agreed that the performance on the indicators were an accurate representation of quality performance and distinguished performance among health plans.

2004 Field Test: Performance Rates on the Systemic Corticosteroid Indicator by Product Line, Age and Gender*

	Denom.	Total Num.	Total Perf. Rate	Active Steroid Rx		New Steroid Rx	
				Num.	Perf. Rate	Num.	Perf. Rate
Product Line							
Commercial	1,085	510	47.0%	127	11.7%	434	40.0%
Medicaid	690	292	42.3%	98	14.2%	234	33.9%
Medicare	693	246	35.5%	61	8.8%	211	30.4%
Age							
40-54	601	244	40.6%	66	11.0%	204	33.9%
55-64	782	394	50.4%	108	13.8%	336	43.0%
65-74	612	258	42.2%	65	10.6%	221	36.1%
75-84	374	125	33.4%	35	9.4%	102	27.3%
85+	99	27	27.3%	12	12.1%	16	16.2%
Gender							
F	1382	602	43.6%	165	11.9%	507	36.7%
M	1086	446	41.1%	121	11.1%	372	34.3%
Total	2468	1048	42.5%	286	11.6%	879	35.6%

**Includes data submitted by 5 Commercial plans, 3 Medicare plans and 1 Medicaid plan using measurement year 2003*

2004 Field Test: Performance Rates on the Bronchodilator Indicator by Product Line, Age and Gender*

	Denom.	Total Num.	Total Perf. Rate	Active Bronch. Rx		New Bronch. Rx	
				Num.	Perf. Rate	Num.	Perf. Rate
Product Line							
Commercial	1,085	571	52.6%	303	27.9%	437	40.3%
Medicaid	690	440	63.8%	275	39.9%	309	44.8%
Medicare	693	312	45.0%	178	25.7%	231	33.3%
Age							
40-54	601	307	51.1%	163	27.1%	238	39.6%
55-64	782	473	60.5%	272	34.8%	346	44.2%
65-74	612	333	54.4%	197	32.2%	241	39.4%
75-84	374	172	46.0%	103	27.5%	124	33.2%
85+	99	38	38.4%	21	21.2%	28	28.3%
Gender							
F	1382	751	54.3%	432	31.3%	538	38.9%
M	1086	572	52.7%	324	29.8%	439	40.4%
Total	2468	1323	53.6%	756	30.6%	977	39.6%

**Includes data submitted by 5 Commercial plans, 3 Medicare plans and 1 Medicaid plan using measurement year 2003*

RESULTS OF CRITICAL DATA ELEMENT VALIDTY: Across four plans, validation of a COPD exacerbation in the medical record was 49%, with a range of 36% to 71%. The health plans were instructed to review hospital records in addition to primary care records to confirm an exacerbation, but feedback from the majority of the plans indicated that chart abstractors were not able to review hospital records and relied on primary care records to note confirmations. Since hospital records are known to be more reliable in documenting care provided in the ED/hospital, this may explain the moderate rate of denominator validation.

2004 Field Test: COPD Exacerbation Medical Record Validation by Plan and Product Line*

	Count of COPD exacerbations confirmed by administrative data	% of patients that had documentation of a COPD exacerbation in medical record	% of patients that did not have documentation of a COPD exacerbation in medical record
Plan:			
A	137	35.8%	51.1%
B	41	46.0%	48.8%
C	51	70.8%	27.5%
D	140	53.1%	43.6%
Total	369	49.1%	44.7%
Product Line:			
Commercial	240	48.8%	43.8%
Medicare	129	49.6%	46.5%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

In four plans, there was 64.2% data consistency for steroid use between administrative and medical record data. This was calculated by adding the percent of steroid use data found in administrative data and medical record data plus the percent of steroids found in neither data source. Of note, a higher percent of steroid use data noted in medical record was not captured via administrative data. This may be due to written prescriptions for steroids that were never filled by the patient. There was 66.6% consistency for bronchodilator use between administrative and medical record data .

2004 Field Test: Systemic Corticosteroid Indicator Validation by Plan*

Plan Code	# of patients with a COPD exacerbation confirmed in both admin & medical record data	% of patients with corticosteroid confirmed in both medical record & admin data	% of patients with corticosteroid confirmed in neither admin or medical record data	% of patients with corticosteroid confirmed in admin data only	% of patients with corticosteroid confirmed in medical record data only
A	39	56.4%	17.9%	5.1%	20.5%
B	17	23.5%	29.4%	0.0%	47.1%
C	34	29.4%	14.7%	5.9%	50.0%
D	69	40.6%	30.4%	10.1%	18.8%
Total	159	40.3%	23.9%	6.9%	28.9%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

2004 Field Test: Bronchodilator Indicator Validation by Plan*

Plan Code	# of patients with a COPD exacerbation confirmed in both admin & medical record data	% of patients with bronchodilator confirmed in both medical record & admin data	% of patients with bronchodilator confirmed in neither admin or medical record data	% of patients with bronchodilator confirmed in admin data only	% of patients with bronchodilator confirmed in medical record data only
A	39	64.1%	0.0%	2.6%	33.3%
B	17	35.3%	29.4%	5.9%	29.4%
C	34	50.0%	2.9%	2.9%	44.1%
D	69	52.2%	23.2%	11.6%	13.0%
Total	159	52.8%	13.8%	6.9%	26.4%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

RESULTS OF CONSTRUCT VALIDITY: The results indicated that the COPD measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized. The level of correlation for Medicaid plans on the spirometry measure and the pharmacotherapy management for COPD exacerbations systemic corticosteroids indicator was moderate (the correlation coefficient was 0.3), while the other correlations were weaker.

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Commercial Plans)*

	Pearson Correlation Coefficients		
	Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator	Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	Spirometry Measure
Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator	-	0.5	0.2
Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	-	-	0.2

**Includes data submitted by 241 Commercial plans to HEDIS for these measures for measurement year 2014*
Note: All correlations are significant at p<.05

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Medicare Plans)*

	Pearson Correlation Coefficients		
	Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator	Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	Spirometry Measure
Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator	-	0.6	0.2
Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	-	-	0.1

**Includes data submitted by 355 Medicare plans to HEDIS for these measures for measurement year 2014*
Note: All correlations are significant at p<.05

Results of Pearson Correlation Coefficient on HEDIS 2015 Asthma Measures (Medicaid Plans)*

	Pearson Correlation Coefficients		
	Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator	Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	Spirometry Measure
Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator	-	0.9	0.3
Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator	-	-	0.2

**Includes data submitted by 124 Medicaid plans to HEDIS for these measures for measurement year 2014*
Note: All correlations are significant at p<.05

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

SYSTEMATIC ASSESSMENT OF FACE VALIDITY: The pharmacotherapy management of COPD exacerbation measure was deemed to have the desirable attributes of a HEDIS measure in 2005 (relevance, scientific soundness, and feasibility). These results indicate the technical expert panels showed good agreement that the measure as specified will accurately differentiate quality across providers. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2005. Our interpretation is that this measure has sufficient face validity.

CRITICAL DATA ELEMENT VALIDTY: The results of the critical data element validity testing demonstrate that the administrative data elements used to calculate the measure denominator (patients who had a COPD exacerbation managed in the emergency department or hospital and were discharged home) and numerator (patients who had a systemic corticosteroid or bronchodilator prescription) had moderate to strong agreement with medical record data and are valid.

CONSTRUCT VALIDITY: Coefficients with absolute value of less than 0.2 are generally considered indicative of weak associations whereas absolute values of 0.2 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that the COPD measures are correlated with each other, suggesting they represent the same underlying quality construct of COPD quality of care.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

N/A

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

N/A

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

N/A

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed

to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model <u>or</u> stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to <u>2b4.9</u>

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*<u>not required</u>, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and

another plan in the 75th percentile of performance. We used these two plans as examples of measured entities. However the method can be used for comparison of any two measured entities.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined)

HEDIS 2014 Variation in Performance across Health Plans: Systemic Corticosteroids Indicator

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Comm. HMO	119	75.4	6.9	67.8	72.1	76.1	79.6	83.7	7.5	0.002
Comm. PPO	184	72.9	7.2	64.3	68.7	73.8	78.2	81.6	9.5	0.002
Medicare HMO	386	71.8	9.0	61.3	69.0	73.5	77.1	80.7	8.1	<0.001
Medicare PPO	463	72.5	10.1	64.7	69.8	73.7	77.8	81.7	8.0	<0.001
Medicaid	497	65.4	13.9	47.6	58.7	69.0	74.6	78.2	15.9	<0.001

EP: Eligible Population, the average denominator size across all plans submitting 2014 HEDIS data for this measure
IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

HEDIS 2014 Variation in Performance across Health Plans: Bronchodilator Indicator

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Comm. HMO	119	80.7	6.6	73.5	76.6	81.0	85.0	89.2	8.4	0.004
Comm. PPO	184	77.6	6.7	68.3	73.5	78.0	81.8	85.3	8.3	0.004
Medicare HMO	386	80.9	7.9	72.2	77.9	81.9	85.7	90.0	7.8	<0.001
Medicare PPO	463	77.6	7.8	68.3	74.9	78.2	82.2	87.2	7.3	<0.001
Medicaid	497	79.0	12.9	64.1	76.1	83.5	87.1	89.0	11.0	<0.001

EP: Eligible Population, the average denominator size across all plans submitting 2014 HEDIS data for this measure
IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
The results above indicate there is a 7-16% gap in performance between the 25th and 75th percentile performing plans across the different product lines and indicators. For all product lines and indicators, the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance is in Medicaid plans, which shows a 16 percentage point gap between 25th and 75th percentile plans for the systemic corticosteroid indicator and an 11 percentage point gap between plans for the bronchodilator indicator.

To put these meaningful differences in performance into context, we estimated that on average 79 additional members per Medicaid plan would have been discharged on a systemic corticosteroid and 55 additional members would have been discharged on a bronchodilator if plans in the 25th percentile performed as well as plans in the 75th percentile. This estimate is based on the average health plan eligible population.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
Plans collect this measure using all administrative data sources. NCQA's audit process verifies that plans' measure calculations are not biased due to missing data.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

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 all 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 followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will 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

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Rating http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Payment Program Medicare Advantage Plan Rating https://www.medicare.gov/find-a-plan/questions/home.aspx NCQA Health Plan Accreditation http://www.ncqa.org/tabid/123/Default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual State of Health Care Quality

	http://www.ncqa.org/tabid/836/Default.aspx Quality Compass http://www.ncqa.org/tabid/177/Default.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.
HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.
HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. [REPLACE or ADD as appropriate, 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives.] Health plans are scored based on performance compared to benchmarks.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
MEDICARE ADVANTAGE PLAN RATING: This measure is included in the composite Medicare Advantage Star Rating. CMS calculates a Star Rating (1-5) for all Medicare Advantage health plans based on 53 performance measures. Medicare beneficiaries can view the star rating and individual measure scores on the CMS Plan Compare website. The Star Rating is also used to calculate bonus payments to health plans with excellent performance. The Medicare Advantage Plan Rating program covers 11.5 million Medicare beneficiaries in 455 health plans across all 50 states.

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

From 2012-2014, the average rate on the systemic corticosteroid indicator showed improvement across Commercial PPO and Medicare HMO and PPO plans (see section 1b.2 for summary of data from health plans). There was also improvement in plans at the 90th percentile for Commercial, Medicare and Medicaid plans on the systemic corticosteroid indicator. These data are nationally representative.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
More Medicaid plans reported the measure in 2014 compared to 2013 and 2012, which may help explain why the average performance rates did not substantially improve. There is hope that with increasing attention to this measure in public reporting programs such as the Medicare Advantage Plan Ratings, performance rates on both indicators will improve.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no identified unintended consequences for this measure during testing or since implementation.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0091 : COPD: Spirometry Evaluation
0102 : COPD: inhaled bronchodilator therapy
0577 : Use of Spirometry Testing in the Assessment and Diagnosis of COPD

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
N/A

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
0091 and 0577 are measures assessing spirometry testing in COPD patients. There is no impact on interpretability or added burden of data collection because the focus of our proposed measure is different. 0102 is a physician-level measure and the focus of our proposed measure is different. Our measure focuses exclusively on patients who were hospitalized or had an ED visit for a COPD exacerbation and received timely recommended treatment (systemic corticosteroids and bronchodilators) while 0102 focuses on managing COPD and allows receipt of a bronchodilator at least once during the measurement year.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955--

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Respiratory Measurement Advisory Panel (RMAP) Members:
David Au, MD, MS, (CHAIR) Associate Prof. of Medicine/Investigator HSRD, Department of Veterans Affairs
Kurt Elward, MD, MPH, Clinical Professor of Family Medicine, Virginia Commonwealth University
Laura Feemster, MD, MS, Assistant Professor of Medicine, University of Washington Medical Center, VA HSR&D
Anne Fuhlbrigge, MD, Clinical Director, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School
Min Joo, MD, MPH, FCCP, Assistant Professor of Medicine, University of Illinois at Chicago/ Jesse Brown VA Medical Center
Christine Joseph, PhD, MPH, BSc, Associate Director of Research, Epidemiologist, Henry Ford Health System
Jerry Krishnan, MD, PhD, Prof. of Medicine & Public Health, Director of Population Health Sciences, AVP, Office of the VP for Health Affairs, University of Illinois Hospital & Health Sciences System
Todd Lee, PharmD, PhD, Primary: Senior Investigator, Secondary: Associate Professor, University of Illinois at Chicago
Allan Luskin, MD, Physician Pulmonologist, Healthy Airways
Richard O´Connor, MD, Director, Dept. of Quality Management, Allergist/Immunologist, Sharp Rees-Stealy Medical Group

Committee on Performance Measurement (CPM) Members:
Peter Bach, MD, Director, Health Policy Center, Memorial Sloan Kettering Cancer Center
Bruce Bagley, MD, FAAFP, Senior Advisor to the Professional Satisfaction and Practice Sustainability effort at the American Medical Association
Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna
Patrick Conway, MD, MSC, Chief Medical Officer and Deputy Administrator, Centers for Medicare and Medicaid Services
Jonathan D. Darer, MD, MPH, Chief Innovation Officer, Geisinger Health System
Helen Darling, Strategic Advisor, National Business Group of Health
Rebekah Gee, MD, MPH, FACOG, Assistant Professor, LSUHSC
Foster Gesten, MD, FACP, New York State Department of Health

Marge Ginsburg, Executive Director, Center for Healthcare Decisions
David Grossman, MD, MPH, Medical Director, Population Health
Christine S. Hunter, MD (Co- Chair), Chief Medical Officer, US Office of Personnel Management
Jeffery Kelman, MMSc, MD, Chief Medical Officer, United Stated Department of Health and Human Services
Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group
J. Brent Pawlecki, MD, MMM, Chief Health Officer, The Goodyear Tire & Rubber Company
Susan Reinhard, PhD, RN, Senior Vice President, AARP Public Policy Institute
Eric C. Schneider, MD, MSc, FACP (Co-chair), Senior Vice President, Policy and Research, The Commonwealth Fund
Marcus Thygeson, MD, MPH, Chief Health Officer, Blue Shield of California

HEDIS EXPERT CODING PANEL Members:
Glen Braden, MBA, CHCA, Attest Health Care Advisors, LLC
Denene Harper, RHIA, American Hospital Association
DeHandro Hayden, BS, American Medical Association
Patience Hoag, RHIT, CPHQ, CHCA, CCS, CCS-P, Health Services Advisory Group
Nelly Leon-Chisen, RHIA, American Hospital Association
Tammy Marshall, LVN, Aetna
Alec McLure, RHIA, CCS-P, Verisk Health
Michele Mouradian, RN, BSN, McKesson Health Solutions
Craig Thacker, RN, CIGNA HealthCare
Mary Jane F. Toomey, RN CPC, Aetna Better Health

Technical Measurement Advisory Panel Members:
Andy Amster, MSPH, Kaiser Permanente
Kathryn Coltin, MPH, Independent Consultant
Lekisha Daniel-Robinson, Centers for Medicare and Medicaid Services
Marissa Finn, MBA, Cigna HealthCare
Scott Fox, MS, MEd, Independence Blue Cross
Carlos Hernandez, CenCalHealth
Kelly Isom, MA, RN, Aetna
Harmon Jordan, ScD, RTI International
Ernest Moy, MD, MPH, Agency for Healthcare Research and Quality
Patrick Roohan, New York State Department of Health
Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC
Natan Szapiro, Independence Blue Cross

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2005
Ad.3 **Month and Year of most recent revision:** 07, 2015
Ad.4 **What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.
Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.
THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a