NATIONAL QUALITY FORUM

Measure Evaluation 4.1
January 2010

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The sub-criteria and most of the footnotes from the evaluation criteria are provided in Word comments and will appear if your cursor is over the highlighted area (or in the margin if your Word program is set to show revisions in balloons). Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized): Complete all yellow highlighted areas of the form. Evaluate the extent to which each sub-criterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: If there is no TAP or workgroup, the SC also evaluates the sub-criteria (yellow highlighted areas).

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the sub-criterion, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed but demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few sub-criteria as indicated)

(for NQF staff use) NQF Review #: ACP-015-10
NQF Project: Ambulatory Care - Additional Outpatient Measures 2010

MEASURE DESCRIPTIVE INFORMATION

De.1 Measure Title: Otitis Media with Effusion: Systemic antimicrobials - Avoidance of inappropriate use
De.2 Brief description of measure: Percentage of patients aged 2 months through 12 years with a diagnosis of OME who were not prescribed systemic antimicrobials
De.3 Type of Measure: process
De.4 National Priority Partners Priority Area: Overuse
De.5 IOM Quality Domain: effectiveness, efficiency, equity
De.6 Consumer Care Need: Getting Better

CONDITIONS FOR CONSIDERATION BY NQF

Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:

A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.
A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)? Yes
A.2 Indicate if Proprietary Measure (as defined in measure steward agreement):
A.3 Measure Steward Agreement: agreement signed and submitted
A.4 Measure Steward Agreement attached:

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section

C. The intended use of the measure includes both public reporting and quality improvement.

Purpose: public reporting, quality improvement

D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement.

D.1 Testing: No, testing will be completed within 12 months

D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures? Yes

(For NQF staff use) Have all conditions for consideration been met?

Staff Notes to Steward (if submission returned):

Met

Staff Notes to Reviewers (issues or questions regarding any criteria):

StaffReviewerName(s):

1. IMPORTANCE TO MEASURE AND REPORT

1a. High Impact

1a.1 Demonstrated High Impact Aspect of Healthcare: affects large numbers

1a.2

1a.3 Summary of Evidence of High Impact: “About 2.2 million diagnosed episodes of OME occur annually in the United States, yielding a combined direct and indirect annual cost estimate of $4.0 billion... About 90% of children (80% of individual ears) have OME at some time before school age, most often between ages 6 months and 4 years. In the first year of life, more than 50% of children will experience OME, increasing to more than 60% by age 2 years. Many episodes resolve spontaneously within 3 months, but about 30% to 40% of children have recurrent OME and 5% to 10% of episodes last 1 year or longer.”


1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: OME usually resolves spontaneously with indications for therapy only if the condition is persistent and clinically significant benefits can be achieved. Antimicrobials have no proven long-term effectiveness and have potential adverse effects.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across...
1c.8 Citations for Evidence (other than guidelines):


<table>
<thead>
<tr>
<th>1c.2-3. Type of Evidence: evidence based guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c.4 Summary of Evidence (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome):</td>
</tr>
<tr>
<td>&quot;Long-term benefits of antimicrobial therapy for OME are unproved despite a modest short-term benefit for 2 to 8 weeks in randomized trials. Initial benefits, however, can become nonsignificant within 2 weeks of stopping the medication. Moreover, about 7 children would need to be treated with antimicrobials to achieve one short-term response. Adverse effects of antimicrobials are significant and may include rashes, vomiting, diarrhea, allergic reactions, alteration of the child’s nasopharyngeal flora, development of bacterial resistance, and cost.&quot;</td>
</tr>
</tbody>
</table>

1c.5 Rating of strength/quality of evidence (also provide narrative description of the rating and by whom):

<table>
<thead>
<tr>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade A</td>
</tr>
</tbody>
</table>

1c.6 Method for rating evidence: "Evidence quality for grades of evidence

Grade A: Well-designed randomized controlled trials or diagnostic studies performed on a population similar to the guideline’s target population

Grade B: Randomized controlled trials or diagnostic studies with minor limitations; overwhelmingly consistent evidence from observational studies

Grade C: Observational studies (case control and cohort design)

Grade D: Expert opinion, case reports, reasoning from first principles (bench research or animal studies)

Grade X: Exceptional situations where validating studies cannot be performed and there is a clear preponderance of benefit over harm

1c.7 Summary of Controversy/Contradictory Evidence:

1c.8 Citations for Evidence (other than guidelines):
**1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number):** Antimicrobials and corticosteroids do not have long-term efficacy and are not recommended for routine management. (Recommendation based on systematic review of randomized, controlled trials and preponderance of harm over benefit. [Aggregate evidence quality - Grade A]) (AAFP/AAO-HNSF/AAP)


**1c.11 National Guideline Clearinghouse or other URL:** http://www.guideline.gov/summary/summary.aspx?doc_id=9310&nbr=004979&string=AAO-HNSF

**1c.12 Rating of strength of recommendation (also provide narrative description of the rating and by whom):**

**Recommendation**

**1c.13 Method for rating strength of recommendation (if different from USPSTF system, also describe rating and how it relates to USPSTF):**

Strong recommendation - A strong recommendation means the benefits of the recommended approach clearly exceed the harms (or that the harms clearly exceed the benefits in the case of a strong negative recommendation) and that the quality of the supporting evidence is excellent (Grade A or B)*. In some clearly identified circumstances, strong recommendations may be made based on lesser evidence when high-quality evidence is impossible to obtain and the anticipated benefits strongly outweigh the harms. Implication: Clinicians should follow a strong recommendation unless a clear and compelling rationale for an alternative approach is present.

Recommendation - A recommendation means the benefits exceed the harms (or that the harms clearly exceed the benefits in the case of a negative recommendation), but the quality of evidence is not as strong (Grade B or C)*. In some clearly identified circumstances, recommendations may be made based on lesser evidence when high-quality evidence is impossible to obtain and the anticipated benefits strongly outweigh the harms. Implication: Clinicians should follow a recommendation unless a clear and compelling rationale for an alternative approach is present.

Option - An option means that either the quality of evidence that exists is suspect (Grade D)* or that well-done studies (Grade A, B, or C)* show little clear advantage to one approach versus another. Implication: Clinicians should be flexible in their decision making regarding appropriate practice, although they may set bounds on alternatives; patient preference should have a substantial influencing role.

No recommendation - No recommendation means there is both a lack of pertinent evidence (Grade D)* and an unclear balance between benefits and harms. Implication: Clinicians should feel little constraint in their decision making and be alert to new published evidence that clarifies the balance of benefit versus harm; patient preference should have a substantial influencing role.

**1c.14 Rationale for using this guideline over others:**

It is the PCPI policy to use guidelines, which are evidence-based, applicable to physicians and other healthcare 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 the quality of care.

**TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Importance to Measure and Report?**

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Rationale</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES**

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable 4
## 2a. MEASURE SPECIFICATIONS

### S.1 Do you have a web page where current detailed measure specifications can be obtained?

#### 2a. Precisely Specified

<table>
<thead>
<tr>
<th>Spec</th>
<th>Rating</th>
</tr>
</thead>
</table>
| 2a.1 Numerator Statement | C

**Comment [KP8]:** The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF’s Health Information Technology Expert Panel (HITEP).

### 2a.2 Numerator Time Window

- **Numerator Time Window:** Once within the denominator time window

### 2a.3 Numerator Details

- **Numerator Details:** EHR specifications for this measure are under development

#### Claims Specifications

- **CPT Category II code:** 4132F - Systemic antimicrobial therapy not prescribed

### 2a.4 Denominator Statement

- **Denominator Statement:** All patients aged 2 months through 12 years with a diagnosis of OME

#### 2a.5 Target population gender:

- **Male, Female**

#### 2a.6 Target population age range:

- **Aged 2 months through 12 years**

#### 2a.7 Denominator Time Window

- **Denominator Time Window:** Each episode* of OME occurring during the twelve consecutive months

**Footnote:** *An episode of OME is defined as a 90-day period from onset of Otitis Media with Effusion (as indicated by the first occurrence of qualifying diagnosis and CPT codes).

### 2a.8 Denominator Details

- **Denominator Details:** EHR specifications for this measure are under development

#### Claims Specifications

- **ICD-9 diagnosis codes:** 381.10, 381.19, 381.20, 381.29, 381.3, 381.4

**AND**

- **CPT codes:** 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99241, 99242, 99243, 99244, 99245, 99381, 99382, 99383, 99384, 99391, 99392, 99393, 99394

### 2a.9 Denominator Exclusions

- **Denominator Exclusions:** Documentation of medical reason(s) for prescribing systemic antimicrobials (e.g., salvage therapy prior to surgery)

#### 2a.10 Denominator Exclusion Details

- **Denominator Exclusion Details:** EHR specifications for this measure are under development

#### Claims Specifications
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2a.11 Stratification Details/Variables</strong></td>
<td>(All information required to stratify the measure including the stratification variables, all codes, logic, and definitions): Stratification by insurance coverage (commercial, Medicare and Medicaid) is recommended by some implementers.</td>
</tr>
<tr>
<td><strong>2a.12-13 Risk Adjustment Type</strong></td>
<td>no risk adjustment necessary</td>
</tr>
<tr>
<td><strong>2a.14 Risk Adjustment Methodology/Variables</strong></td>
<td>(List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method):</td>
</tr>
<tr>
<td><strong>2a.15-17 Detailed risk model available</strong></td>
<td>Web page URL or attachment:</td>
</tr>
<tr>
<td><strong>2a.18-19 Type of Score</strong></td>
<td>rate/proportion</td>
</tr>
<tr>
<td><strong>2a.20 Interpretation of Score</strong></td>
<td>better quality = lower score</td>
</tr>
<tr>
<td><strong>2a.21 Calculation Algorithm</strong></td>
<td>(Describe the calculation of the measure as a flowchart or series of steps): See sample calculation algorithm attached</td>
</tr>
<tr>
<td><strong>2a.22 Describe the method for discriminating performance</strong></td>
<td>(e.g., significance testing):</td>
</tr>
<tr>
<td><strong>2a.23 Sampling (Survey) Methodology</strong></td>
<td>If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):</td>
</tr>
<tr>
<td><strong>2a.24 Data Source</strong></td>
<td>(Check the source(s) for which the measure is specified and tested) Electronic administrative data/claims, electronic Health/Medical Record, paper medical record/flowsheet, special or unique data</td>
</tr>
<tr>
<td><strong>2a.25 Data source/data collection instrument</strong></td>
<td>(Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.):</td>
</tr>
<tr>
<td><strong>2a.26-28 Data source/data collection instrument reference web page URL or attachment:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>2a.29-31 Data dictionary/code table web page URL or attachment:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>2a.32-35 Level of Measurement/Analysis</strong></td>
<td>(Check the level(s) for which the measure is specified and tested) Clinicians: Individual, Clinicians: Group</td>
</tr>
<tr>
<td><strong>2a.36-37 Care Settings</strong></td>
<td>(Check the setting(s) for which the measure is specified and tested) Ambulatory Care: Office, Ambulatory Care: Clinic, Ambulatory Care: Hospital Outpatient</td>
</tr>
<tr>
<td><strong>2a.38-41 Clinical Services</strong></td>
<td>(Healthcare services being measured, check all that apply) Clinicians: PA/NP/Advanced Practice Nurse, Clinicians: Physicians (MD/DO)</td>
</tr>
</tbody>
</table>

**TESTING/ANALYSIS**

### 2b. Reliability testing

**2b.1 Data/sample** (description of data/sample and size): This measure is utilized in the QuIlN project sponsored Care for Children with Otitis Media with Effusion Project. This AHRQ funded study is in progress, at the Center for Health Care Quality at the Cincinnati Children’s Hospital Medical Center. Participating physicians pull 30 charts of otitis media patients with effusion and review the performance measures for feasibility, also assessing inter-rater reliability. The project will use simple frequencies and the kappa statistic as a measure of inter-rater reliability to assess feasibility and reliability of data collection.

**Comment [KP10]:** 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.
2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
The PCPI supports the consideration of exceptions (or exclusions) on a measure by measure basis. There must be a clear rationale to permit an exception for a medical, patient, or system reason, based on whether or not that reason is significant and occurs frequently enough. The PCPI also advocates for the systematic review and analysis of each physician’s exceptions data to identify practice patterns and opportunities for quality improvement. That is, while exceptions are removed from the denominator when calculating performance, rates of exceptions should be reported alongside performance rates.

Denominator exceptions are included in this particular measure so that physicians can identify patients for whom systemic antimicrobials may be appropriate (e.g., salvage therapy prior to surgery, patients with a coexisting condition that warrants the use of these agents).

2d.2 Citations for Evidence:

2d.3 Data/sample (description of data/sample and size):

2d.4 Analytic Method (type analysis & rationale):

2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses):

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample (description of data/sample and size):

2e.2 Analytic Method (type of risk adjustment, analysis, & rationale):

2e.3 Testing Results (risk model performance metrics):

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

2b.2 Analytic Method (type of reliability & rationale, method for testing):

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

2c. Validity testing

2c.1 Data/sample (description of data/sample and size):

2c.2 Analytic Method (type of validity & rationale, method for testing):

It is the consensus of the PCPI Measures Implementation and Evaluation Committee that face and content validity of PCPI measures can be assumed to be established once they have progressed beyond the Public Comment period by virtue of the specialized expertise of the PCPI work group members who are involved in identifying and drafting performance measures within a topical domain as well, as the rigorous, structured discussions that are prescribed according to PCPI protocols for work group conduct.

2c.3 Testing Results (statistical results, assessment of adequacy in the context of norms for the test conducted):

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample (description of data/sample and size):

2e.2 Analytic Method (type of risk adjustment, analysis, & rationale):

2e.3 Testing Results (risk model performance metrics):

QuINN Project Overview and Requirements. Available at: http://www.aap.org/qualityimprovement/quin/OME.html.

Comment [k11]: 8 Examples of reliability testing 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 may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity assessed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be: supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; AND a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus; AND precisely defined and specified: ... [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated: an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care; clinical judgment not defined. OR ... [3]

Comment [k17]: 12 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, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by rr ... [4]
2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use (description of data/sample and size):

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (type of analysis & rationale):

2f.3 Provide Measure Scores from Testing or Current Use (description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance):

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample (description of data/sample and size):

2g.2 Analytic Method (type of analysis & rationale):

2g.3 Testing Results (e.g., correlation statistics, comparison of rankings):

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results (scores by stratified categories/cohorts):

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:

TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Scientific Acceptability of Measure Properties?

Steering Committee: Overall, to what extent was the criterion, Scientific Acceptability of Measure Properties, met? Rationale:

3. USABILITY

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

3a. Meaningful, Understandable, and Useful Information

3a.1 Current Use: testing not yet completed

3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years):

This measure is used in the CMS PQRI program claims option for 2008.

3a.3 If used in other programs/initiatives (If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years):

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [K19]: 14 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% v. 75%) 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 poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender); OR rationale/data justifies why stratification is not necessary or not feasible.

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus groups, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.
Testing of Interpretability  (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)

3a.4 Data/sample (description of data/sample and size):

3a.5 Methods (e.g., focus group, survey, QI project):

3a.6 Results (qualitative and/or quantitative results and conclusions):

3b/3c. Relation to other NQF-endorsed measures

3b.1 NQF # and Title of similar or related measures:

(for NQF staff use) Notes on similar/related endorsed or submitted measures:

<table>
<thead>
<tr>
<th>3b</th>
<th>C</th>
<th>P</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population):

3b.2 Are the measure specifications harmonized? If not, why?

3c. Distinctive or Additive Value
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:

5.1 Competing Measures If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), describe why it is a more valid or efficient way to measure quality:

TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Usability?

3c | C | P | M | N |

4. FEASIBILITY

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

<table>
<thead>
<tr>
<th>4a</th>
<th>C</th>
<th>P</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Generated as a Byproduct of Care Processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4a.1-2 How are the data elements that are needed to compute measure scores generated? data generated as byproduct of care processes during delivery, coding/abstraction performed by someone other than person obtaining original information,

4b. Electronic Sources

4b.1 Are all the data elements available electronically? (elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims)

No

4b.2 If not, specify the near-term path to achieve electronic capture by most providers. Electronic health record products are not in ability to collect data in a standardized way at this

| 4b | C | P | M | N |

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
4c. Exclusions

4c.1. Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?
No

4c.2. If yes, provide justification.

4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences

4d.1. Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.
Physicians have voluntarily reported on this measure as part of the PQRI program. We are not aware of any unintended consequences related to this measurement.

4e. Data Collection Strategy/Implementation

4e.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/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/implementation issues:

4e.2. Costs to implement the measure (costs of data collection, fees associated with proprietary measures):

4e.3. Evidence for costs:

4e.4. Business case documentation:

TAP/Workgroup: What are the strengths and weaknesses in relation to the sub-criteria for Feasibility?

Steering Committee: Overall, to what extent was the criterion, Feasibility, met?
Rationale:

RECOMMENDATION

(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.

Steering Committee: Do you recommend for endorsement?
Comments:

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
American Medical Association | 515 N State St. | Chicago | Illinois | 60654

Co.2 Point of Contact
Mark | Antman, DDS, MBA | mark.antman@ama-assn.org | 312-464-5056

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
**Measure Developer if different from Measure Steward**

<table>
<thead>
<tr>
<th>Co.3 Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Medical Association</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co.4 Point of Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co.5 Submitter if different from Measure Steward POC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co.6 Additional organizations that sponsored/participated in measure development</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Academy of Otolaryngology - Head and Neck Surgery (AAO-HNS) Foundation</td>
</tr>
</tbody>
</table>

## ADDITIONAL INFORMATION

**Workgroup/Expert Panel involved in measure development**

Ad.1 Provide a list of sponsoring organizations and workgroup/panel members’ names and organizations.

Describe the members’ role in measure development.

- Allan S. Lieberthal, MD, FAAP (Co-Chair) (pediatrics)
- Richard M. Rosenfeld, MD, MPH (Co-Chair) (otolaryngology)
- Brian L. Bachelder, MD (family medicine)
- Steve I. Pelton, MD (pediatrics/pediatric infectious diseases)
- Karen Jo Doyle, MD, PhD (otolaryngology)
- Peter S. Roland, MD (otolaryngology)
- Cynthia P. Helstad, PhD, RN
- Xavier Sevilla, MD (pediatrics)
- Rahul Khare, MD, FACEP (emergency medicine)
- David L. Witsell, MD, MHS (otolaryngology)

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 must be 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.

Ad.2 If adapted, provide name of original measure:
Ad.3-5 If adapted, provide original specifications URL or attachment

**Measure Developer/Steward Updates and Ongoing Maintenance**

Ad.6 Year the measure was first released: 2007
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure? Every 3 years or as new evidence becomes available that materially affects the measures
Ad.9 When is the next scheduled review/update for this measure? 2010-03

Ad.10 Copyright statement/disclaimers: Physician Performance Measures (Measures) and related data specifications, developed by the Physician Consortium for Performance Improvement® (the Consortium), are intended to facilitate quality improvement activities by physicians.

These Measures are intended to assist physicians in enhancing quality of care. Measures are designed for use by any physician who manages the care of a patient for a specific condition or for prevention. These performance Measures are not clinical guidelines and do not establish a standard of medical care. The Consortium has not tested its Measures for all potential applications. The Consortium encourages the testing and evaluation of its Measures.

Measures are subject to review and may be revised or rescinded at any time by the Consortium. The Measures may not be altered without the prior written approval of the Consortium. Measures developed by the Consortium, 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 American Medical Association, on behalf of the Consortium.

Neither the Consortium nor its members shall be responsible for any use of these Measures.

THE MEASURES ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND

© 2007 American Medical Association. All Rights Reserved

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. The AMA, the Consortium and its members disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

THE SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.11 -13 Additional Information web page URL or attachment: Attachment Sample Calculation Algorithm-6340070722593772.doc

Date of Submission (MM/DD/YY): 02/17/2010
1c. The measure focus is:
- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed; OR
- if an intermediate outcome, process, structure, etc., there is evidence that supports the specific measure focus as follows:
  - Intermediate outcome - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process - evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure - evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience - evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency - demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

2d. Clinically necessary measure exclusions are identified and must be:
- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; AND
- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus; AND
- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
  - if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and 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).

2e. For outcome measures and other measures (e.g., resource use) when indicated:
- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care; OR
- rationale/data support no risk adjustment.

13 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, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.
Sample PCPI Calculation Algorithm

Calculation for Performance
For performance purposes, a measure is calculated by creating a fraction with the following components: Numerator, Denominator, and Denominator Exclusions.

Numerator (A) Includes:
Number of patients meeting numerator criteria

Denominator (PD) Includes:
Number of patients meeting criteria for denominator inclusion

Denominator Exclusions (C) Include:
Number of patients with valid medical, patient or system exclusions (where applicable; will differ by measure)

Performance Calculation

\[
\frac{A}{PD - C}
\]

If a measure does not allow for exclusion(s), it is calculated by creating a fraction with the following components: Numerator and Denominator.

Numerator (A) Includes:
Number of patients meeting numerator criteria

Denominator (PD) Includes:
Number of patients meeting criteria for denominator inclusion


\[
\frac{A}{PD}
\]

It is also possible to calculate the percentage of patients excluded overall, or excluded by medical, patient, or system reason where applicable:

Overall Exclusion Calculation

\[
\frac{C}{PD}
\]

OR

Exclusion Calculation by Type

\[
\frac{C_1}{PD}
\]

\[
\frac{C_2}{PD}
\]

\[
\frac{C_3}{PD}
\]