**NATIONAL QUALITY FORUM**

*Measure Submission and Evaluation Worksheet 5.0*

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

<table>
<thead>
<tr>
<th>NQF #: 0096</th>
<th>NQF Project: Pulmonary Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>(for Endorsement Maintenance Review)</td>
<td></td>
</tr>
<tr>
<td>Original Endorsement Date: May 01, 2007</td>
<td>Most Recent Endorsement Date: Jul 31, 2012</td>
</tr>
</tbody>
</table>

**BRIEF MEASURE INFORMATION**

- **De.1 Measure Title:** Empiric Antibiotic for Community-Acquired Bacterial Pneumonia
- **Co.1.1 Measure Steward:** American Medical Association - Physician Consortium for Performance Improvement (AMA-PCPI)
- **De.2 Brief Description of Measure:** Percentage of patients aged 18 years and older with a diagnosis of community-acquired bacterial pneumonia with an appropriate empiric antibiotic prescribed
- **2a1.1 Numerator Statement:** Patients with appropriate empiric antibiotic prescribed
- **2a1.4 Denominator Statement:** All patients aged 18 years and older with a diagnosis of community-acquired bacterial pneumonia
- **2a1.8 Denominator Exclusions:**
  - Documentation of medical reason(s) for not prescribing appropriate empiric antibiotic
  - Documentation of patient reason(s) for not prescribing appropriate empiric antibiotic
  - Documentation of system reason(s) for not prescribing appropriate empiric antibiotic

- **1.1 Measure Type:** Process
- **2a1. 25-26 Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Pharmacy, Electronic Clinical Data : Registry, Paper Medical Records
- **2a1.33 Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

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

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

**STAFF NOTES (issues or questions regarding any criteria)**

**Comments on Conditions for Consideration:**

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

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

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

**Other Criteria:**

**Staff Reviewer Name(s):**

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

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

Created on: 10/04/2012 at 02:44 PM
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.

1a. High Impact: [H☐ M☐ L☐ I☐]
(The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.)

De.4 Subject/Topic Areas (Check all the areas that apply): Pulmonary/Critical Care, Pulmonary/Critical Care: Pneumonia
De.5 Cross Cutting Areas (Check all the areas that apply):

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

1a.2 If “Other,” please describe:

1a.3 Summary of Evidence of High Impact (Provide epidemiologic or resource use data):
Community-acquired pneumonia (CAP) is a common condition and a leading cause of morbidity and mortality. The bacterial agent streptococcus pneumoniae is the most common etiology of community-acquired pneumonia in outpatients and hospitalized patients including those in the ICU. (1) Other bacterial causes of CAP include nontypeable Haemophilus influenzae and Moraxella catarrhalis, generally in patients who have underlying bronchopulmonary disease, and S. aureus, especially during an influenza outbreak. (1)

Pneumonia and Influenza are the eighth leading cause of mortality in the United States, together accounting for approximately 56,000 deaths in 2008. (2) More than 54,000 or 96% of these deaths are attributed solely to pneumonia. (2)

CAP has a significant economic impact. Estimates suggest that the annual cost of treating CAP in the United States is $12.2 billion. (3) In 2007, CAP resulted in approximately 1.1 million hospitalizations and accounted for an estimated 4.5 million ambulatory care visits, including 32,000 visits to emergency departments. (4,5)

Approximately 10% of patients hospitalized for CAP are admitted to the intensive care unit. (6) CAP in the ICU is associated with a high mortality rate of about 30%. (7)

Additionally, CAP is one of the two most common reasons for potentially preventable hospitalizations, accounting for an estimated $7.2 billion of all preventable hospitalizations. (8)

admitted to the ward and the ICU. Chest. 2008; 133:610-617

1b. Opportunity for Improvement: H□ □ M□ □ L□ □ I □
(There is a demonstrated performance gap - variability or overall less than optimal performance)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
Current diagnostic methods to determine CAP pathogens are limited. Guidelines state that infections with the overwhelming majority of CAP pathogens will be adequately treated by use of the recommended empirical regimens. Additionally, rapid and appropriate empirical antibiotic therapy is consistently associated with improved outcome (IDSA/ATS guideline)

1b.2 Summary of Data Demonstrating Performance Gap (Variation or overall less than optimal performance across providers):
[For Maintenance – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.]
Current evidence suggests variation in the prescription of empiric antibiotic for patients with community-acquired bacterial pneumonia. In one retrospective cohort study of 631 patients, 357 (57%) received guideline-concordant empiric antibiotic therapy.(1)

In another study of 54,619 non–intensive care unit inpatients with CAP, 35,477 (65%) received initial guideline-concordant therapy. After adjustment for severity of illness and other confounders, guideline-concordant therapy was associated with improved health outcomes and diminished resource use in adults. These findings were linked to treatment with fluoroquinolone or macrolide agents.(2)

CMS Physician Quality Reporting Initiative/System:
This measure was used in the CMS Physician Quality Reporting Initiative/System (PQRI/S) in the 2007 through 2010 claims option as well as the registry and measure group options for 2009 and 2010. There is a gap in care as shown by this 2008 data; 22.52% of patients reported on did not meet the measure.(3)

10th percentile: 33.33 %
25th percentile: 66.67 %
50th percentile: 90.91 %
75th percentile: 100.00 %
90th percentile: 100.00%
Exception rate: 0.79%

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

1b.4 Summary of Data on Disparities by Population Group: [For Maintenance – Descriptive statistics for performance results for this measure by population group]
Data support the existence of racial, gender and age disparities in CAP prevention, treatment and mortality.

Prevention
Vaccination to prevent pneumonia is associated with improved survival, decreased chance of respiratory failure or other complications, and decreased length of stay among hospitalized patients with community
acquired pneumonia.(1)

According to the CDC, more than 34% of adults aged 65 and older reported not ever receiving a pneumococcal vaccination. Racial disparities exist in the receipt of pneumococcal vaccine. Whites have the highest rate of pneumococcal vaccination at 67.8% while Blacks and Hispanics have lower rates of 52.5% and 51.3%, respectively.(2)

Treatment
Community-acquired bacterial pneumonia disproportionately affects the elderly. Individuals aged 65 and over are hospitalized at a rate that is almost 6 times higher as that of individuals aged 45-64 and about 23 times higher than those aged 15-44.(3)

Mortality
According to the American Lung Association, there are racial and gender disparities in the mortality rate due to pneumonia and influenza. In 2006, Black men were approximately 16% more likely to die from pneumonia and influenza than White men.(3)

Males tend to have higher death rates due to pneumonia than females. In 2006, the age adjusted death rates for females and males were 15.5 and 21.2 per 100,000, respectively.(4)

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

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


Does the measure pass subcriterion1c?
M-H □ M-H □ □ □ □ Yes □
L □ □ □ □ □ □ □ Yes □ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No □
M-H □ □ □ □ □ □ □ Yes □ IF potential benefits to patients clearly outweigh potential harms: otherwise No □
L-M-H □ □ □ □ □ □ □ No □

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service

Does the measure pass subcriterion1c?
Yes □ IF rationale supports relationship

1c.1 Structure-Process-Outcome Relationship (Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process-health outcome; intermediate clinical outcome-health outcome): The measure focuses on the prescription of empiric antibiotic for community-acquired bacterial pneumonia. The current limitations of diagnostic methods for community-acquired bacterial pneumonia result in the prescription of empiric antibiotic therapy for most patients with CAP. Infections with the overwhelming majority of CAP pathogens will be adequately treated by use of the recommended empirical regimens. Additionally, rapid and appropriate empirical antibiotic therapy is consistently associated with improved outcome. (IDSA/ATS guideline)
1c.2-3 Type of Evidence (Check all that apply):
Clinical Practice Guideline

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 measure focus is on adults 18 years and older with a diagnosis of community-acquired bacterial pneumonia. According to the IDSA/ATS 2007 guidelines, the bacterial agent streptococcus pneumoniae is the most common etiology of community-acquired pneumonia in outpatients and hospitalized patients including those in the ICU.

The prescription of appropriate empiric antibiotic is recommended for this population. The evidence cited for this measure is directly related to the limitations of current CAP diagnostic methods which result in the need for empiric antibiotic therapy in most patients with CAP even if organism-directed antimicrobial therapy would be ideal.

1c.5 Quantity of Studies in the Body of Evidence (Total number of studies, not articles): The ATS guidelines cited nine studies in support of the recommendation.

The IDSA 2003 guidelines do not address the quantity of studies supporting the recommendation but cited a total of 235 references.

The IDSA/ATS 2007 guidelines also do not address the quantity of studies supporting their recommendation but cited a total of 335 references.

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 quality of the evidence supporting the guideline recommendations was not comprehensively addressed but described and graded as follows:
The ATS 2001 guidelines recommend that patients be treated empirically with advanced generation macrolide: azithromycin or clarithromycin or doxycycline if no cardiopulmonary disease and no modifying factors exist, or with beta lactam plus macrolide or doxycycline or fluoroquinolone alone if cardiopulmonary disease and/or risk factors are present. The quality of the evidence supporting the ATS 2001 guideline recommendation is not comprehensively addressed but is described as Level II evidence.

The IDSA 2003 guidelines recommend initial empiric therapy of selected outpatients with CAP. The evidence supporting this recommendation is described as Level A Recommendation - Level I Evidence. Fluoroquinolones (gatifloxacin, gemifloxacin, levofloxacin, and moxifloxacin) are recommended for initial empiric therapy of selected outpatients with CAP (Level A Recommendation - Level I Evidence). Other options (macrolides and doxycycline) are generally preferred for uncomplicated infections in outpatients (Level A Recommendation - Level I Evidence). A macrolide is recommended as monotherapy for selected outpatients, such as those who were previously well and not recently treated with antibiotics (Level A Recommendation - Level I Evidence). A macrolide plus a beta-lactam is recommended for initial empiric treatment of outpatients in whom resistance is an issue and for hospitalized patients (Level A Recommendation - Level I Evidence).

The IDSA/ATS 2007 guidelines recommend a macrolide (azithromycin, clarithromycin, or erythromycin) for outpatients who were previously healthy with no risk factors for drug-resistant S. pneumoniae. The evidence supporting this recommendation is described as being Strong recommendation; Level I evidence. The guidelines also recommend Doxycycline and the supporting evidence is described as Weak recommendation; Level III evidence.

1c.7 Consistency of Results across Studies (Summarize the consistency of the magnitude and direction of the effect): Although there is no explicit statement regarding the overall consistency of results across studies, the IDSA/ATS 2007 guidelines state that the only randomized controlled trial of diagnostic strategy in CAP has demonstrated no statistically significant differences in mortality rate or LOS between patients receiving pathogen directed therapy and patients receiving empirical therapy.
The development of the ATS 2001 guidelines was by a committee composed of pulmonary, critical care, infectious disease and general internal medicine specialists, in an effort to incorporate a variety of perspectives. The guideline committee’s process to develop the guideline and recommendations is described and based on deliberations that culminated in committee consensus that the recommendation is appropriate.

The development of the IDSA/ATS 2007 guidelines was by a committee that consisted of infectious diseases, pulmonary, and critical care physicians with interest and expertise in pulmonary infections.

The development of the IDSA 2003 guidelines was not addressed.

1c.8 Net Benefit (Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms):
Current limitations of diagnostic methods for community-acquired bacterial pneumonia result in a reliance on empiric antibiotic therapy in most patients with CAP. Infections with the overwhelming majority of CAP pathogens will be adequately treated by use of the recommended empirical regimens. Additionally, rapid and appropriate empirical antibiotic therapy is consistently associated with improved outcome. (IDSA/ATS guideline)

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 development of the ATS 2001 guidelines was by a committee composed of pulmonary, critical care, infectious disease and general internal medicine specialists, in an effort to incorporate a variety of perspectives and to create a statement that was acceptable to a wide range of physicians. The committee graded the evidence supporting their recommendations.

The committee that developed the IDSA 2003 guidelines used the Infectious Diseases Society of America–United States Public Health Service grading system for rating recommendations in clinical guidelines. The committee provided the following disclosures: Lionel A. Mandell has received research funding from Bayer, Bristol-Myers Squibb, and Pharmacia; has been a consultant for Bayer, Pfizer, Aventis, Ortho-McNeil, and Janssen-Ortho; and has been on the speakers’ bureau for Pfizer, Aventis,Wyeth, Ortho-McNeil, and Bayer.
Thomas M. File, Jr., has received research funding from Abbott, AstraZeneca, Bayer, Bristol-Myers Squibb, Cubist, GlaxoSmithKline, Pfizer, and Wyeth; has been a consultant for Aventis, Bayer, Cubist, GlaxoSmithKline, Ortho-McNeil, Pfizer, and Wyeth; and has been on the speakers’ bureau for Abbott, Aventis, Bayer, GlaxoSmithKline, Merck, Ortho-McNeil, Pfizer, and Wyeth.

The development of the IDSA/ATS 2007 guidelines was by a committee that consisted of infectious diseases, pulmonary, and critical care physicians with interest and expertise in pulmonary infections. This committee also graded the evidence supporting its recommendations and provided the following disclosures:
L.A.M. has received research funding from Bayer, Chiron, Ortho-McNeil, Oscient, and Pfizer; has served as a consultant to Bayer, Cempra, Novexel, Ortho-McNeil, Oscient, Pfizer, Sanofi-Aventis, Targanta, and Wyeth; and has served on speakers’ bureaus for Bayer, Ortho-McNeil, Oscient, Pfizer, and Sanofi-Aventis. R.G.W. has received research funding from Chiron, Eli Lilly, Pfizer, andWyeth; has served on the Clinical Evaluation Committee for Johnson and Johnson; has served as a clinical trial participant in studies initiated by Takeda, Biosite, Inverness Medical Intervention, Johnson and Johnson, and Altana; and has served as consultant to the Oklahoma Foundation for Medical Quality and the Centers for Medicare and Medicaid Services. J.G.B. serves on the advisory board of Johnson and Johnson.
T.M.F. has received research funding from Binax Incorporated, Ortho-McNeil, Oscient, Pfizer, and Sanofi Aventis; has served as a consultant to Bayer, GlaxoSmithKline, Merck, Ortho-McNeil, Oscient, Pfizer, Sanofi-Aventis, Schering-Plough, andWyeth; and has served on speakers’ bureaus for Abbott, GlaxoSmithKline, Merck, Ortho-McNeil, Oscient, Pfizer, Sanofi-Aventis, Schering-Plough, and Wyeth. N.A.D has received research support from Altana and Sanofi-Aventis; has served on the advisory boards for Sanofi-Aventis and AstraZeneca; and has served on the speakers’ bureaus for Pfizer, Schering-Plough, Sanofi-Aventis, and Merck. A.A. has served on the speakers’ bureaus for Altana, Bayer Pharma, Boehringer-Ingehelm, Chiron, Elan, GlaxoSmithKline, Ortho-McNeil, Pfizer, and Sanofi-Aventis; has served as a consultant and on advisory boards for Altana, Bayer Pharma, Boehringer-
NQF #0096 Empiric Antibiotic for Community-Acquired Bacterial Pneumonia, Last Updated Date: Sep 10, 2012

Ingelheim, Chiron, Elan, GlaxoSmithKline, Ortho-McNeil, Pfizer, and Sanofi-Aventis; and has received research funding from BART, Bayer Pharma, Boehringer-Ingelheim, GlaxoSmithKline, and Lilly.
M.S.N. serves on the speakers' bureaus for and as a consultant to AstraZeneca, Aventis, Elan, Merck, Ortho-McNeil, Pfizer, Schering-Plough, and Wyeth.
All other authors: no conflicts.

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

1c.12 If other, identify and describe the grading scale with definitions: In grading the evidence supporting the guideline recommendations, the ATS 2001 and the IDSA/ATS 2007 guidelines committees used the following scale:
Level I (high) Evidence from well-conducted, randomized controlled trials.
Level II (moderate) Evidence from well-designed, controlled trials without randomization (including cohort, patient series, and case-control studies); also include any large case series in which systematic analysis of disease patterns and/or microbial etiology was conducted, as well as reports of data on new therapies that were not collected in a randomized fashion.
Level III (low) Evidence from case studies and expert opinion. In some instances, therapy recommendations come from antibiotic susceptibility data without clinical observations.

The IDSA 2003 committee used the following Infectious Diseases Society of America–United States Public Health Service grading system:

Strength of recommendation
A Good evidence to support a recommendation for use
B Moderate evidence to support a recommendation for use
C Poor evidence to support a recommendation
D Moderate evidence to support a recommendation against use
E Good evidence to support a recommendation against use

Quality of evidence
I Evidence from > or = 1 properly randomized, controlled trial
II Evidence from > or = 1 well-designed clinical trial, without randomization; from cohort or case-controlled analytic studies (preferably from >1 center); from multiple time-series; or from dramatic results of uncontrolled experiments
III Evidence from opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

1c.13 Grade Assigned to the Body of Evidence: Same as above Section 1C.6. Quality of Body of Evidence

1c.14 Summary of Controversy/Contradictory Evidence: The IDSA/ATS 2007 guideline states that the only randomized controlled trial of diagnostic strategy in CAP demonstrated that pathogen-directed therapy was associated with lower mortality among the small number of patients admitted to the ICU. The study was performed in a country with a low incidence of antibiotic resistance, which may limit its applicability to areas with higher levels of resistance. Adverse effects were significantly more common in the empirical therapy group but may have been unique to the specific antibiotic choice (erythromycin). The guideline concludes that the lack of benefit overall in this trial should not be interpreted as a lack of benefit for an individual patient.

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

1c.16 Quote verbatim, the specific guideline recommendation (including guideline # and/or page #): ATS 2001 guidelines
Empiric therapy based on the initial CAP guidelines leads to a better outcome than if nonguideline therapy is used (Level II evidence). page 1742
Outpatients with no history of cardiopulmonary disease, and no modifying factors should be treated with advanced generation macrolide: azithromycin or clarithromycin or Doxycycline.
Outpatients with cardiopulmonary disease, and/or other modifying factors should be treated with a Beta-Lactam (oral cefpodoxime, cefuroxime, high-dose amoxicillin, amoxicillin/clavulanate; or parenteral ceftriaxone followed by oral cefpodoxime) plus Macrolide or
doxycycline or antipneumococcal fluoroquinolone (used alone). Tables 2 and 3; pages 1735-1736.

IDSA 2003 guidelines
Fluoroquinolones (gatifloxacin, gemifloxacin, levofloxacin, and moxifloxacin) are recommended for initial empiric therapy of selected outpatients with CAP (A-I). Other options (macrolides and doxycycline) are generally preferred for uncomplicated infections in outpatients (A-I). A macrolide is recommended as monotherapy for selected outpatients, such as those who were previously well and not recently treated with antibiotics (A-I). A macrolide plus a beta-lactam is recommended for initial empiric treatment of outpatients in whom resistance is an issue and for hospitalized patients (A-I). Table 4 page 1413

IDSA/ATS 2007 guidelines
Outpatient treatment
1. For previously healthy and no use of antimicrobials within the previous 3 months: macrolide (strong recommendation; level I evidence); Doxycycline (weak recommendation; level III evidence)
2. Presence of comorbidities such as chronic heart, lung, liver or renal disease; diabetes mellitus; alcoholism; malignancies; asplenia; immunosuppressing conditions or use of immunosuppressing drugs; or use of antimicrobials within the previous 3 months: respiratory fluoroquinolone (moxifloxacin, gemifloxacin, or levofloxacin [750 mg]) (strong recommendation; level I evidence); Beta-lactam plus a macrolide (strong recommendation; level I evidence). Table 7 page S45

1c.17 Clinical Practice Guideline Citation: ATS 2001 guidelines


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: Same as above section 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 grading the evidence supporting the guideline recommendations, the ATS 2001 and the IDSA/ATS 2007 guidelines committees used the following scale: 
Level I (high) Evidence from well-conducted, randomized controlled trials.
Level II (moderate) Evidence from well-designed, controlled trials without randomization (including cohort, patient series, and case-control studies); also include any large case series in which systematic analysis of disease patterns and/or microbial etiology was conducted, as well as reports of data on new therapies that were not collected in a randomized fashion.
Level III (low) Evidence from case studies and expert opinion. In some instances, therapy recommendations come from antibiotic
susceptibility data without clinical observations.

In grading the strength of the evidence supporting the guideline recommendation, the IDSA/ATS 2007 guideline committee also described the following:

Strong recommendation is that most patients should receive that intervention. Conversely, moderate or weak recommendations suggest that, even if a majority would follow the recommended management, many practitioners may not.

For the final document, a strong recommendation required >= 6 (of 12) of the members to consider it to be strong and the majority of the others to grade it as moderate.

The IDSA 2003 guidelines committee used the following Infectious Diseases Society of America–United States Public Health Service grading system:

Strength of recommendation
A Good evidence to support a recommendation for use
B Moderate evidence to support a recommendation for use
C Poor evidence to support a recommendation
D Moderate evidence to support a recommendation against use
E Good evidence to support a recommendation against use

Quality of evidence
I Evidence from > or = 1 properly randomized, controlled trial
II Evidence from > or = 1 well-designed clinical trial, without randomization; from cohort or case-controlled analytic studies (preferably from >1 center); from multiple time-series; or from dramatic results of uncontrolled experiments
III Evidence from opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

1c.23 Grade Assigned to the Recommendation: Same as above Section 1C.6. Quality of Body of Evidence

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: Moderate 1c.26 Quality: High 1c.27 Consistency: Moderate
1c.28 Attach evidence submission form:
1c.29 Attach appendix for supplemental materials:

Was the threshold criterion, Importance to Measure and Report, met? (1a & 1b must be rated moderate or high and 1c yes) Yes No

Provide rationale based on specific subcriteria:

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

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

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

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.
S.1 Measure Web Page (In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained). Do you have a web page where current detailed specifications for this measure can be obtained? Yes

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

<table>
<thead>
<tr>
<th>2a. RELIABILITY. Precise Specifications and Reliability Testing:</th>
<th>H □ M □ L □ I □</th>
</tr>
</thead>
</table>

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

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

2a1.2 Numerator Time Window (The time period in which the target process, condition, event, or outcome is eligible for inclusion):
Once for each episode of CAP during measurement period

2a1.3 Numerator Details (All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses):
Numerator Instructions:
This measure is to be reported once for each occurrence of community-acquired bacterial pneumonia during the reporting period. Each unique occurrence is defined as a 45-day period from onset of community-acquired bacterial pneumonia.

Numerator Definitions:
Appropriate Empiric Antibiotic – For treatment of community-acquired bacterial pneumonia (CAP) should include any medication from one of the following four drug classes: Fluoroquinolones, Macrolides, Doxycycline, Beta Lactam with Macrolide or Doxycycline (classes as defined by current ATS/IDSA guidelines; antibiotics within these classes and FDA-approved for outpatient CAP treatment may be considered).
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.

For EHR:
eSpecification currently under development. Data elements (using Quality Data Model) required for the measure attached.

For Claims/Administrative:
CPT Category II code: 4045F: Appropriate empiric antibiotic prescribed

2a1.4 Denominator Statement (Brief, narrative description of the target population being measured):
All patients aged 18 years and older with a diagnosis of community-acquired bacterial pneumonia

2a1.5 Target Population Category (Check all the populations for which the measure is specified and tested if any): Adult/Elderly Care

2a1.6 Denominator Time Window (The time period in which cases are eligible for inclusion):
Each episode of CAP during 12 month measurement period

2a1.7 Denominator Details (All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):
For EHR:
eSpecification currently under development. Data elements (using Quality Data Model) required for the measure attached.

For Claims/Administrative:
Patients aged >= 18 years on date of encounter AND
Denominator Exclusions

2a1.8 Denominator Exclusions (Brief narrative description of exclusions from the target population):
- Documentation of medical reason(s) for not prescribing appropriate empiric antibiotic
- Documentation of patient reason(s) for not prescribing appropriate empiric antibiotic
- Documentation of system reason(s) for not prescribing appropriate empiric antibiotic

2a1.9 Denominator Exclusion Details (All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):
The PCPI methodology uses three categories of reasons for which a patient may be excluded from the denominator of an individual measure. 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 measure Empiric Antibiotic for Community-Acquired Bacterial Pneumonia, exceptions may include medical reasons, patient reasons or system reasons. Where examples of exceptions are included in the measure language, these examples are coded and included in the eSpecifications. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients’ medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician’s exceptions data to identify practice patterns and opportunities for quality improvement. For example, it is possible for implementers to calculate the percentage of patients that physicians have identified as meeting the criteria for exception. Additional details by data source are as follows:

For EHR:
eSpecification currently under development. Data elements (using Quality Data Model) required for the measure attached.

For Claims/Administrative:
- Documentation of medical reason(s) for not prescribing appropriate empiric antibiotic - Append modifier to CPT Category II code: 4045F-1P
- Documentation of patient reason(s) for not prescribing appropriate empiric antibiotic - Append modifier to CPT Category II code: 4045F-2P
- Documentation of system reason(s) for not prescribing appropriate empiric antibiotic - Append modifier to CPT Category II code: 4045F-3P

2a1.10 Stratification Details/Variables (All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses):
We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

2a1.11 Risk Adjustment Type (Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13):
- No risk adjustment or risk stratification
- 2a1.12 If "Other," please describe:
2a1.13 Statistical Risk Model and Variables (Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.): None

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

2a1.17-18. Type of Score: Rate/proportion

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

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

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. If the measure does not have exceptions, STOP. If the measure has exceptions, proceed with the following steps: 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 the 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.

Calculation algorithm is included in data dictionary/code table attachment 2a1.30.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment: Attachment
AMA-PCPI_Measure Calculation-Standard Measures-634625099605483392.pdf

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):
Not applicable. The measure does not require sampling or a survey

2a1.25 Data Source (Check all the sources for which the measure is specified and tested). If other, please describe:
Administrative claims, Electronic Clinical Data, Electronic Clinical Data: Electronic Health Record, Electronic Clinical Data: Pharmacy, Electronic Clinical Data: Registry, Paper Medical Records

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable
Created on: 10/04/2012 at 02:44 PM
2a1.26 **Data Source/Data Collection Instrument** *(Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.): Not applicable*

2a1.27-29 **Data Source/data Collection Instrument Reference Web Page URL or Attachment:**

2a1.30-32 **Data Dictionary/Code Table Web Page URL or Attachment:**
Attachment
AMA-PCPI_0096_EM-CAP_EmpAntibio_DE_08282012.pdf

2a1.33 **Level of Analysis** *(Check the levels of analysis for which the measure is specified and tested): Clinician : Group/Practice, Clinician : Individual, Clinician : Team*

2a1.34-35 **Care Setting** *(Check all the settings for which the measure is specified and tested): Ambulatory Care : Clinician Office/Clinic, Ambulatory Care : Urgent Care, Home Health, Hospital/Acute Care Facility, Other:Emergency Department, 'Domiciliary, Rest Home or Custodial Care Services', Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility*

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

2a2.1 **Data/Sample** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included): Rush University Medical Center Testing Project
The data source was Electronic Medical Records in an Emergency Room setting
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 100 patient encounters.
Data collected on patients seen between January 1 – December 31, 2009
Visual inspection of the medical record was performed in 4Q2010

Brigham and Women’s Hospital CAP Measure Testing
The data source was Electronic Health Records at 8 Primary Care clinics
The patient population consisted of adults with an administrative claims diagnosis of pneumonia during the measurement period
The reviewers agreed that 198 encounters to 71 different clinicians were visits for acute pneumonia to comprise the sample population

University of Chicago Testing Project
The data source was Paper Medical Records in an Emergency Room setting
Trained abstractors reviewed a sample 151 medical records

2a2.2 **Analytic Method** *(Describe method of reliability testing & rationale): Data abstracted from randomly sampled patient records were used to calculate inter-rater reliability for the measure.

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

Rush University Medical Center Testing Project
Randomly select 100 patients from the eligible population
Reviewer 1 reviewed patients 1-75 for numerator, denominator, exceptions
Reviewer 2 reviewed patients 26-100 for numerator, denominator, exceptions
Calculated inter-rater reliability on patients 26-75
Calculated parallel forms reliability on patients 1-100 against automated report

Brigham and Women’s Hospital CAP Measure Testing
Retrospective, cross-sectional electronic chart review
Two trained reviewers independently abstracted charts
Inter-rater reliability was calculated

University of Chicago Testing Project
A paper abstraction tool was developed from the measure specifications
A total of 151 charts were reviewed by trained research nurse abstractors
Data from the paper abstraction tools were entered into a database
Performance was calculated
Inter-rater reliability was calculated using a Kappa statistic

2a2.3 Testing Results (Reliability statistics, assessment of adequacy in the context of norms for the test conducted):
Rush University Medical Center Testing Reliability: N, % Agreement, Kappa, (95% CI)
This measure demonstrates perfect agreement.
Overall Reliability: 100, 100%, kappa non-calculable*

Brigham and Women’s Hospital Testing Reliability: Kappa, (95% CI)
This measure demonstrates substantial to almost perfect agreement.
Overall Reliability: any antibiotic 0.85 (0.70 – 0.99)
Overall Reliability: appropriate antibiotic 0.76 (0.66 – 0.86)

University of Chicago Testing Reliability: N, Kappa, (95% CI)
This measure demonstrates almost perfect agreement.
Overall Reliability: Antibiotic given in ER: 110, 0.93 (0.87 – 1.00)
Overall Reliability: Patient discharged home: 109, 0.95 (0.89 – 1.00)
Overall Reliability: Patient prescribed antibiotic: 69, 0.90 (0.83– 0.95)

* Kappa statistics cannot be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

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:
The measure focus is the process of prescribing empiric antibiotic to patients 18 years of age and over with community-acquired bacterial pneumonia (CAP). The measure is specified for treatment of CAP with any medication from one of the following four drug classes: Fluoroquinolones, Macrolides, Doxycycline, Beta Lactam with Macrolide or Doxycycline (classes as defined by current ATS/IDSA guidelines; antibiotics within these classes and FDA-approved for outpatient CAP treatment may be considered). 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.

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):
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

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable
Created on: 10/04/2012 at 02:44 PM
14
Rush University Medical Center Testing Project
The data source was Electronic Medical Records in an Emergency Room setting.
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 100 patient encounters.
Data collected on patients seen between January 1 – December 31, 2009
Visual inspection of the medical record was performed in 4Q2010

Face Validity
An expert panel was used to assess face validity of the measure. This panel consisted of 23 members, with representation from a number of specialties including critical care, emergency, family, geriatrics, internal medicine and radiology.

Bruce S. Auerbach, MD, FACEP (Co-Chair)
Eric C. Schneider, MD, MSc (Co-Chair)
James G. Adams, MD, FACEP
Dennis M. Beck, MD, FACEP
Raj Behal, MD, MPH
Stephen V. Cantrill, MD, FACEP
Randall B. Case, MD, FACEP
William C. Dalsey, MD, FACEP
Andrew Eisenberg, MD, MHA
Robert Emmick, Jr., MD, FACEP
James Feldman, MD, MPH
Paul Gitman, MD, MACP
Richard T. Griffey, MD, MPH
Scott R. Gunn, MD
Steven D. Hanks, MD, MMM, FACP
Jeffrey P. Kanne, MD
Rahul Khare, MD
Sravanthi Reddy, MD
Carlotta M. Rinke, MD, FACP, MBA
Sam JW Romeo, MD, MBA
John F. Schneider, MD PhD
John. J. Skendzielewski, MD, FACEP
Carl Tommaso, MD FSCAI

2b2.2 Analytic Method (Describe method of validity testing and rationale; if face validity, describe systematic assessment):
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) and the measure overall.
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

Face Validity
All PCPI performance measures are assessed for content validity by expert Work Group members during the development process. Additional input on the content validity of draft measures is obtained through a 30-day public comment period and by also soliciting comments from a panel of consumer, purchaser, and patient representatives convened by the PCPI specifically for this purpose. All
comments received are reviewed by the expert Work Group and the measures adjusted as needed. Other external review groups (i.e. focus groups) may be convened if there are any remaining concerns related to the content validity of the measures.

An expert panel was used to assess face validity of the measure. This panel consists of 23 members, with representation from a number of specialties including critical care, emergency, family, geriatrics, internal medicine and radiology.

The aforementioned panel was asked to rate their agreement with the following statement:
The scores obtained from the measure as specified will accurately differentiate quality across providers.

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

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):
EHR Measure Validity
This measure demonstrates moderate agreement when comparing the automated EHR report to visual inspection.

Rush University Medical Center Testing Project: N, % Agreement, Kappa, (95% CI)
Numerator Reliability: 100, 82.9%, 0.56 (0.37-0.75)
Denominator Reliability: 100, 72.2%, Kappa non-calculable*
Overall Reliability: 100, 82.9%, 0.56 (0.37-0.75)

* Kappa statistics cannot be calculated because there are a substantial proportion of zeros. 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 = 14; Mean rating = 4.21 and 85.7% 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 - 1
3 - 1 (Neither Disagree nor Agree)
4 - 6
5 - 6 (Strongly Agree)

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):
Rush University Medical Center Testing Project
The data source was Electronic Medical Records in an Emergency Room setting
The data sample came from 1 site representing an academic medical center located in an urban area.
The sample consisted of 100 patient encounters.
Data collected on patients seen between January 1 – December 31, 2009
Visual inspection of the medical record was performed in 4Q2010

University of Chicago Testing Project
The data source was Paper Medical Records in an Emergency Room setting
Trained abstractors reviewed a sample 151 medical records
2b3.2 **Analytic Method** *(Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference):*

Exceptions included medical, patient and system reasons. Exceptions were analyzed for frequency and variability across providers.

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

Rush University Medical Center Testing Project  
Exception rate: 2.1%

University of Chicago Testing Project  
Exception rate: 6.25%

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):*  
This measure is not risk adjusted

2b4.2 **Analytic Method** *(Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables):*  
This measure is not risk adjusted

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):*  
Not applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:  
As a process measure, no risk adjustment is necessary.

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):*  
CMS Physician Quality Reporting Initiative/System:  
206,209 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: #29 Empiric Antibiotic for Community-acquired Bacterial Pneumonia  
# Eligible Professionals: 203,563  
# Professionals Reporting: 22,677  
% Professionals Reporting: 11.14%  
# Professionals Reporting >=80% of eligible instances: 17,233  
% Professionals Reporting >=80% of eligible instances: 75.99%

2b5.2 **Analytic Method** *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance):*  
CMS Physician Quality Reporting Initiative/System:  
The inter-quartile range (IQR) was calculated to determine the variability of performance on the measure.

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):

CMS Physician Quality Reporting Initiative/System:
Scores on this measure: N = 206,209; Mean = 77.48%
10th percentile: 33.33 %
25th percentile: 66.67 %
50th percentile: 90.91 %
75th percentile: 100.00%
90th percentile: 100.00%

The inter-quartile range (IQR) provides a measure of the dispersion of performance. The IQR is 33.33 and indicates that 50% of physicians have performance on this measure ranging from 66.67% and 100.00% and 10% of physicians have performance rates less than or equal to 33.33%.(1)

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

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):

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

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

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) and the measure overall.
• Kappa statistic to ensure that agreement rates are not a phenomenon of chance

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):

EHR Measure Validity
This measure demonstrates moderate agreement when comparing the automated EHR report to visual inspection.

Rush University Medical Center Testing Project: N, % Agreement, Kappa, (95% CI)
Numerator Reliability: 100, 82.9%, 0.56 (0.37-0.75)
Denominator Reliability: 100, 72.2%, Kappa non-calculable*
Overall Reliability: 100, 82.9%, 0.56 (0.37-0.75)

* Kappa statistics cannot be calculated because there are a substantial proportion of zeros. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

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): We encourage
the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
The PCPI advocates that performance measure data should, where possible, be stratified by race, ethnicity, and primary language to assess disparities and initiate subsequent quality improvement activities addressing identified disparities, consistent with recent national efforts to standardize the collection of race and ethnicity data. A 2008 NQF report endorsed 45 practices including stratification by the aforementioned variables.(1) A 2009 IOM report recommends collection of the existing Office of Management and Budget (OMB) race and Hispanic ethnicity categories as well as more fine-grained categories of ethnicity (referred to as granular ethnicity and based on one’s ancestry) and language need (a rating of spoken English language proficiency of less than very well and one’s preferred language for health-related encounters).”(2)

References:

as well as the registry and measure group options for 2009 and 2010.

The PCPI believes that the reporting of participation information is a beneficial first step on a trajectory toward the public reporting of performance results, which is appropriate since the measure has been tested and the reliability of the performance data has been validated. Continued NQF endorsement will facilitate our ongoing progress toward this public reporting objective.

3a.2 Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: The PCPI believes that the reporting of participation information is a beneficial first step on a trajectory toward the public reporting of performance results, which is appropriate since the measure has been tested and the reliability of the performance data has been validated. Continued NQF endorsement will facilitate our ongoing progress toward this public reporting objective.

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s): This measure may be used in a Maintenance of Certification program.

3b. Usefulness for Quality Improvement: H M L I
(The measure is meaningful, understandable and useful for quality improvement.)

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

All PCPI measures are suitable for use in quality improvement initiatives and are made freely available on the PCPI website and through the implementation efforts of medical specialty societies and other PCPI members. The PCPI strongly encourages the use of its measures in QI initiatives and seeks to provide information on such initiatives to PCPI members.

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated (e.g., QI initiative), describe the data, method and results: The PCPI believes that the use of PCPI measures in quality improvement initiatives is a beneficial way to gather scientific data with which to improve physician performance. This is appropriate since the measure has been tested and the reliability of the performance data has been validated. NQF endorsement will facilitate our ongoing progress toward this quality improvement objective.

Overall, to what extent was the criterion, Usability, met? H M L I
Provide rationale based on specific subcriteria:

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

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

4a.1-2 How are the data elements needed to compute measure scores generated? (Check all that apply).
Data used in the measure are:
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

4b. Electronic Sources: H M L I

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

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

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

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable
Created on: 10/04/2012 at 02:44 PM
4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:
We are not aware of any unintended consequences related to this measurement.

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

A.2 Please check if either of the following apply (regarding proprietary measures):
4d.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues (e.g., fees for use of proprietary measures):
This measure was found to be reliable and feasible for implementation.

Overall, to what extent was the criterion, Feasibility, met? H ☐ M ☐ L ☐ I ☐
Provide rationale based on specific subcriteria:

### OVERALL SUITABILITY FOR ENDORSEMENT

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

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

### 5. COMPARISON TO RELATED AND COMPETING MEASURES

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

5.1 If there are related measures (either same measure focus or target population) or competing measures (both the same measure focus and same target population), list the NQF # and title of all related and/or competing measures:
0147 : Initial antibiotic selection for community-acquired pneumonia (CAP) in immunocompetent patients

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications completely harmonized? No

5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:
Measure #0147 is related to #0096. However, the level of analysis differs; measure #0147 is specified at the facility level and measure #0096 at the individual level. Additionally, measure #0147 relates to the receipt of antibiotic treatment during a hospitalization whereas measure #0096 includes patients with an emergency room discharge diagnosis of community-acquired bacterial pneumonia. Furthermore, we have developed and will maintain specifications for multiple data sources, including Electronic Health Records (EHRs) and Claims-Based Reporting. Our specifications for EHRs are developed in accordance with the terminology standards (eg, SNOMED, RxNorm, LOINC) named in the Meaningful Use Program (CMS EHR Incentive Program).

5b. Competing Measure(s)

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

### CONTACT INFORMATION

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable
Created on: 10/04/2012 at 02:44 PM
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.
Bruce S. Auerbach, MD, FACEP (Co-Chair)
Eric C. Schneider, MD, MSc (Co-Chair)
James G. Adams, MD, FACEP
Dennis M. Beck, MD, FACEP
Raj Behal, MD, MPH
Stephen V. Cantrill, MD, FACEP
Randall B. Case, MD, FACEP
William C. Dalsey, MD, FACEP
Andrew Eisenberg, MD, MHA
Robert Emmick, Jr., MD, FACEP
James Feldman, MD, MPH
Paul Gitman, MD, MACP
Richard T. Griffey, MD, MPH
Scott R. Gunn, MD
Steven D. Hanks, MD, MMM, FACP
Jeffrey P. Kanne, MD
Rahul Khare, MD
Sravanthi Reddy, MD
Carlotta M. Rinke, MD, FACP, MBA
Sam JW Romeo, MD, MBA
John F. Schneider, MD PhD
John. J. Skiendzielewski, MD, FACEP
Carl Tommaso, MD FSCAI

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

<table>
<thead>
<tr>
<th>Measure Developer/Steward Updates and Ongoing Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad.3 Year the measure was first released: 2006</td>
</tr>
<tr>
<td>Ad.4 Month and Year of most recent revision: 12, 2011</td>
</tr>
<tr>
<td>Ad.5 What is your frequency for review/update of this measure? Coding/specification updates occur annually</td>
</tr>
<tr>
<td>Ad.6 When is the next scheduled review/update for this measure? 2012</td>
</tr>
</tbody>
</table>

Ad.7 Copyright statement: Physician Performance Measures (Measures) and related data specifications, developed by the American Medical Association (AMA) in collaboration with the Physician Consortium for Performance Improvement (the Consortium) and the National Committee for Quality Assurance (NCQA) pursuant to government sponsorship under subcontract 6205-05-054 with Mathematica Policy Research, Inc. under contract 500-00-0033 with Centers for Medicare & Medicaid Services.

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, 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 AMA, (on behalf of the Consortium) or NCQA. Neither the AMA, NCQA, the Consortium nor its members shall be responsible for any use of the Measures.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND. © 2004-6 American Medical Association and National Committee for Quality Assurance. 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, NCQA, 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.

Ad.8 Disclaimers: See copyright statement above.

Ad.9 Additional Information/Comments:

**Date of Submission (MM/DD/YY):** 10/18/2011
<table>
<thead>
<tr>
<th>GDM* Standard Category</th>
<th>GDM* Data Type</th>
<th>Standard Terminology (Value Set OID)</th>
<th>Constraints</th>
<th>Value Set Name</th>
<th>Value of Data Element</th>
<th>Data Source</th>
<th>Comments/Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure Timing</td>
<td>N/A</td>
<td>TSB to measure numerator</td>
<td>Measurement Start Date</td>
<td>TSB to measure numerator</td>
<td>Measurement Start Date</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
</tr>
<tr>
<td>Initial Characteristic</td>
<td>Patient Characteristic</td>
<td>Gender (Value Set OID) (2.16.840.1.113883.1.11.1)</td>
<td>during measurement period</td>
<td>Gender</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
<td></td>
</tr>
<tr>
<td>Initial Characteristic</td>
<td>Patient Characteristic</td>
<td>Race (Value Set OID) (2.16.840.1.113883.1.11.2)</td>
<td>during measurement period</td>
<td>Race</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
<td></td>
</tr>
<tr>
<td>Initial Characteristic</td>
<td>Patient Characteristic</td>
<td>Ethnicity (Value Set OID) (2.16.840.1.113883.1.11.3)</td>
<td>during measurement period</td>
<td>Ethnicity</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
<td></td>
</tr>
<tr>
<td>Initial Characteristic</td>
<td>Patient Characteristic</td>
<td>Payer Source of Payment Typology Value Set (TBD)</td>
<td>during measurement period</td>
<td>Payer</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
<td></td>
</tr>
<tr>
<td>Initial Characteristic</td>
<td>Patient Characteristic</td>
<td>Language (Value Set OID) (2.16.840.1.113883.1.11.4)</td>
<td>during measurement period</td>
<td>Preferred Language</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
<td></td>
</tr>
<tr>
<td>Calculated</td>
<td>Age (Value Set OID) (2.16.840.1.113883.3.360.102.4)</td>
<td>during the start of measurement period</td>
<td>Age</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition / Diagnosis / Problem</td>
<td>Diagnosis, Active</td>
<td>Emergency Medicine: Community-Acquired Pneumonia (CAP): Empiric Antibiotic</td>
<td>during measurement period</td>
<td>Emergency Medicine: Community-Acquired Pneumonia (CAP): Empiric Antibiotic</td>
<td>Emergency Medicine: Community-Acquired Pneumonia (CAP): Empiric Antibiotic</td>
<td>Electronic Health Record (EHR)</td>
<td>This data element is collected for the purpose of stratifying results in an effort to highlight disparities.</td>
</tr>
<tr>
<td>Encounter</td>
<td>Encounter, Performed</td>
<td>CPT (2.16.840.1.113883.3.464.0003.01.0005)</td>
<td>during measurement period</td>
<td>Office Visit, Occurrence A</td>
<td>CPT</td>
<td>Electronic Health Record (EHR)</td>
<td>It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. If there is more than one Encounter within 45 days, report the first Encounter during the 45 day episode period.</td>
</tr>
<tr>
<td>Medication</td>
<td>Medication, Active</td>
<td>Fluroquinolones (RxNorm Value Set OID) (TBD)</td>
<td>during measurement period</td>
<td>Fluroquinolones</td>
<td>Medication</td>
<td>Electronic Health Record (EHR)</td>
<td>It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. If there is more than one Encounter within 45 days, report the first Encounter during the 45 day episode period.</td>
</tr>
<tr>
<td>Medication</td>
<td>Medication, Active</td>
<td>Doxycycline (RxNorm Value Set OID) (TBD)</td>
<td>during measurement period</td>
<td>Doxycycline</td>
<td>Medication</td>
<td>Electronic Health Record (EHR)</td>
<td>It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. If there is more than one Encounter within 45 days, report the first Encounter during the 45 day episode period.</td>
</tr>
<tr>
<td>Medication</td>
<td>Medication, Active</td>
<td>Macrolides (RxNorm Value Set OID) (TBD)</td>
<td>during measurement period</td>
<td>Macrolides</td>
<td>Medication</td>
<td>Electronic Health Record (EHR)</td>
<td>It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. If there is more than one Encounter within 45 days, report the first Encounter during the 45 day episode period.</td>
</tr>
<tr>
<td>Medication</td>
<td>Medication, Active</td>
<td>Beta lactam (RxNorm Value Set OID) (TBD)</td>
<td>during measurement period</td>
<td>Beta lactam</td>
<td>Medication</td>
<td>Electronic Health Record (EHR)</td>
<td>It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. It is expected that this numerator action will be performed and reported once for EACH occurrence of CAP. If there is more than one Encounter within 45 days, report the first Encounter during the 45 day episode period.</td>
</tr>
</tbody>
</table>

*The Quality Data Model (GDM), Version 2.1, was developed by National Quality Forum (NQF). ©2011 American Medical Association. All rights reserved.*
### Basic Measure Calculation:

\[
\frac{(N)}{(D) - (E)} = \%
\]

The PCPI strongly recommends that exception rates also be computed and reported alongside performance rates as follows:

### Exception Calculation:

\[
\frac{(E)}{(D)} = \%
\]

### Exception Types:

\(E = E1 \text{ (Medical Exceptions)} + E2 \text{ (Patient Exceptions)} + E3 \text{ (System Exceptions)}\)

For patients who have more than one valid exception, only one exception should be counted when calculating the exception rate.

### Initial Patient Population (IPP)

**Definition:** The initial patient population identifies the general group of patients that the performance measures designed to address; usually focused on a specific clinical condition (e.g., coronary artery disease, asthma). For example, a patient aged 18 years and older with a diagnosis of CAD who has at least 2 visits during the measurement period.

- Find the patients who meet the Initial Patient Population criteria (IPP)

### Denominator (D)

**Definition:** The denominator defines the specific group of patients for inclusion in a specific performance measure based on specific criteria (e.g., patient's age, diagnosis, prior MI). In some cases, the denominator may be identical to the initial patient population.

- Find the patients who qualify for the denominator (D):
  - From the patients within the Patient Population criteria (IPP) select those people who meet Denominator selection criteria.
  - (In some cases the IPP and D are identical).

### Numerator (N)

**Definition:** The numerator defines the group of patients in the denominator for whom a process or outcome of care occurs (e.g., flu vaccine received).

- Find the patients who qualify for the Numerator (N):
  - From the patients within the Denominator (D) criteria, select those people who meet Numerator selection criteria.
  - Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.

### Denominator Exceptions (E)

**Definition:** Denominator exceptions are the valid reasons why patients who are included in the denominator population did not receive a process or outcome of care (described in the numerator). Patients may have Denominator Exceptions for medical reasons (e.g., patient has an egg allergy so they did not receive flu vaccine); patient reasons (e.g., patient declined flu vaccine); or system reasons (e.g., patient did not receive flu Vaccine due to vaccine shortage). These cases are removed from the denominator population for the performance calculation, however the number of patients with valid exceptions should be calculated and reported. This group of patients constitutes the Denominator Exception reporting population – patients for whom the numerator was not achieved and a there is a valid Denominator Exception.

- From the patients who did not meet the Numerator criteria, determine if the patient meets any criteria for the Denominator Exception (E1 + E2 + E3). If they meet any criteria, they should be removed from the Denominator for performance calculation.
  - As a point of reference, these cases are removed from the denominator population for the performance calculation, however the number of patients with valid exceptions should be calculated and reported.