NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple and Blue text represents the responses from measure developers

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0069

Corresponding Measures:

De.2. Measure Title: Appropriate Treatment for Upper Respiratory Infection

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The Appropriate Treatment for Upper Respiratory Infection (URI) measure assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection were not dispensed an antibiotic prescription. The measure includes patients enrolled in commercial, Medicaid, and Medicare health plans.

1b.1. Developer Rationale: This measure assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection (URI) were not dispensed an antibiotic prescription. The improvement in quality envisioned by the use of this measure is to decrease unnecessary prescribing of antibiotic treatment for upper respiratory infection. Too often, antibiotics are prescribed inappropriately, which can lead to antibiotic resistance (when antibiotics can no longer cure bacterial infections).

S.4. Numerator Statement: The numerator of the measure includes the number of dispensed prescriptions for an antibiotic medication on or 3 days after the Episode Date.

S.6. Denominator Statement: Episodes for members 3 months of age and older as of July 1 of the year prior to the measurement year who had an outpatient, telephone, e-visit or virtual check-in, an observation visit or ED encounter with a diagnosis of upper respiratory infection (URI) during the intake period (July 1st of the year prior to the measurement year to June 30th of the measurement year).

S.8. Denominator Exclusions: Exclude visits that result in an inpatient stay.

Exclude Episode Dates when the member had a claim/encounter with any diagnosis for a comorbid condition during the 12 months prior to or on the Episode Date.

Exclude Episode Dates where a new or refill prescription for an antibiotic medication was filled 30 days prior to the Episode Date or was active on the Episode Date.

Exclude Episode Dates where the patient had a claim/encounter with a competing diagnosis on or three days after the Episode Date.

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jan 07, 2013

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

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

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

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

Summary of prior review in 2012

- This is a maintenance process measure utilizing claims data at the health plan level that assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection were not dispensed an antibiotic prescription.
- In its 2012 submission, the developer cited a systematic review titled, "Principles of Appropriate Antibiotic Use: Part II. Nonspecific Upper Respiratory Tract."

Changes to evidence from last review

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

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

Updates:

- The developer provides a logic model depicting a reduction in the inappropriate dispensing of antibiotics for upper respiratory infections, often referred to as "a common cold" could lead to fewer

strains of antibiotic-resistant bacteria and reduction in community-acquired antibiotic resistant infections

- The developer cites two Cochrane systematic reviews and one clinical practice guideline as evidence to support the measurement of antibiotics being prescribed to treat upper respiratory infections.
    - One Cochrane Review found that there is no difference in clinical outcomes for prescription of immediate, delayed or no antibiotics (moderate quality).
    - The other Cochrane Review found that there is no evidence of benefit and there are adverse effects of prescribing antibiotics for children or adults for purulent rhinitis. There is no grade
- The developer cited the clinical practice guideline from the Institute for Clinical Systems called, "Diagnosis and treatment of respiratory illness in children and adults."

Exception to evidence
Not Applicable

Questions for the Committee:

- What is the relationship of this measure to patient outcomes?
- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high (Box 5) → High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

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

- For the current submission, the developer provided the following commercial and Medicaid rates:
    - For 2018 through 2020, the commercial HEDIS data below covers 384, 378, and 404 health plans respectively.
        - 2020: Mean = 78.66%, Std Dev = 8.59%, Min = 48.51%, 10th - 90th Percentile Range = 67.59% - 89.00%, Max = 96.17%, IQR = 11.34%
        - 2019: Mean = 90.06%, Std Dev = 5.97%, Min = 58.42%, 10th - 90th Percentile Range = 82.29% - 95.84%, Max = 100%, IQR = 5.83%
        - 2018: Mean = 88.26%, Std Dev = 7.09%, Min = 44.11%, 10th - 90th Percentile Range = 79.21% - 94.93%, Max = 100%, IQR = 7.03%
    - For 2018 through 2020, the Medicaid HEDIS data below covers 207, 199, and 222 health plans respectively.
        - 2020: Mean = 89.92%, Std Dev = 6.79%, Min = 49.96%, 10th - 90th Percentile Range = 78.79% - 93.53%, Max = 100.00%, IQR = 5.84%
        - 2019: Mean = 90.45%, Std Dev = 6.36%, Min = 65.15%, 10th - 90th Percentile Range = 82.04% - 96.79%, Max = 100.00%, IQR = 6.97%
        - 2018: Mean = 89.08%, Std Dev = 7.11%, Min = 60.25%, 10th - 90th Percentile Range = 80.55% - 95.94%, Max = 100.00%, IQR = 7.14%

Disparities

- The developer noted that HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare), but could be stratified by demographic variables.

- The developer noted that HEDIS does include two measures which can be used to assess disparities in the health plan population.
- The developer noted that there is very little data related to upper respiratory infections and racial and ethnic disparities.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- If no or limited disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- developer provided updated evidence for this measure
- Process measure - This measure assesses whether patients received abx for URI. The goal is to reduce abx use for URIs which are most likely to be viral in nature. This applies directly. I am not aware of new studies.
- Evidence to support (updated) for avoiding use of antibiotics in individuals with upper respiratory tract infections. No difference in between those not treated and those treated. (Cochran Review)
- Evidence is updated and applies directly to process being measured.
- Evidence to support measure focus is sound
- Evidence is strong for the question.
- This is a maintenance process measure for persons 3 months of age and older. Confusion arises in that the title for measure #0069 is "Appropriate Treatment for CHILDREN with Upper Respiratory Infection", while in the submission the title is "Appropriate Treatment for Upper Respiratory Infections". On page 29 of the Measure Worksheet under section S3.2 is found, "The measure was recently expanded to include adults (previously limited to children 3 months to 18 years of age). The measure as currently specified assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection (URI) were not dispensed an antibiotic prescription. NCQA expanded the age range for this measure based on stakeholder and measurement advisory panel feedback ... The measure was also changed from a member-based measure to an episode-based measure. The member based denominator resulted in members with multiple URIs during the measurement period counting only once. An episode-based measure captures more episodes of potentially inappropriate antibiotic prescribing." These are both significant changes to this maintenance measure. The desired outcome is fewer prescriptions written to treat upper respiratory infections. The literature review cited in the recent submission does include studies on children and adults. However, the grade of the evidence in this ICSI review was low.
- Evidence is strong for this measure. The measure developer/steward cited two Cochrane systemic reviews and one clinical practice guideline as evidence to support the measure. One review found that

there is no difference in clinical outcomes for prescription of antibiotics. The other review found no benefits and there were adverse effects of prescribing antibiotics for children or adults. The clinical guideline specified by the measure developer was from the Institute for Clinical Systems. There is a strong relationship to patient outcomes. The evidence provided was directly applicable to the process of care being measured.

- Process measure, evidence rating is high

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- Opportunity for improvement exist
- Although performance in this measure is better than in the Bronchitis measure, there is still a performance gab. in 2020, the mean was ~79% with variability. performance was better in the medicaid data.
- Data provided shows that commercial plans had an increase in antibiotic prescribing for URI in 2020 while Medicaid plans have been steady(~90%) in the past three years. No disparities data was provided due to how data is captured. The disparities question is perhaps important given to a shift to virtual visits in the face of SARS-COV-2.
- Performance data provided that demonstrates a gap.
- yes, improving performance gap year to year, but there is still a gap. 2020 data may be COVID-impacted
- Performance gap is documented. Concern that overarching quality of facility makes a bigger difference than individual provider. Can developer explain what systems are at work in this outcome.
- Claims data were analyzed from health plans for years 2018 to 2020. It is not clear when the measure was changed to include adults as well as children and when the measure was changed from member based to episode based. No target percentage of URI episodes that do not result in antibiotic prescriptions is mentioned. There is not a trend in percentages in the Medicaid group or the Commercial group over the three years. HEDIS data is used. It is mentioned that two HEDIS measures can be used to assess disparities in health plan populations but I did not find that these two measures were applied to the study populations for this measure. The developer noted that there is very little data related to upper respiratory infections and racial and ethnic disparities.
- 's The measure developer's data did not identify a racial disparity. Commercial data means identified variance from 2018 to 2019. Medicaid data was relatively stable for the three measurement years. Higher disparity in commercial versus the Medicaid population in results.
- Evidence suggests a performance gap exists

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

2c. For composite measures: empirical analysis support composite approach

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluators: NQF Staff

Staff Review

Specifications:

- Since last endorsement, the developer expanded the eligible population by broadening the age range and changed the measure to an episode-based measure.

Reliability

- The developer provided measure score reliability and construct validity using data from all health plans that submitted HEDIS data for this measure in 2019. These analyses included 404 Commercial and 222 Medicaid health plans.
- Using 2019 HEDIS data, the developer used a beta-binominal model to assess the signal-to-noise ratio. Using this method, the mean commercial reliability score was 0.983 and the mean Medicaid reliability score was 0.92.
- For signal to noise, the developer states a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The testing suggests that all indicators within this measure have good reliability between 0.7 and 1.0.

Validity:

- Validity testing was performed for the measure score.
    - The developer conducted correlation analysis with three measures.
    - Results:
        - Positive Correlation: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis
            - Medicaid: Correlation coefficient = 0.68, $p < 0.001$
            - Commercial: Correlation coefficient = 0.68, $p < 0.001$
        - Positive Correlation: Use of Imaging Studies for Low Back Pain
            - Medicaid: Correlation coefficient = 0.41, $p < 0.001$
            - Commercial: Correlation coefficient = 0.622, $p < 0.001$
        - Negative Correlation: Antibiotic Utilization
            - Medicaid: Correlation coefficient = -0.73, $p < 0.00$ 1
            - Commercial: Correlation coefficient = -0.74, $p < 0.001$
        - The developer concluded that plans that perform well on this measure are likely to perform well on the positively correlated measures.

- The developer also conducted face validity for this measure. Provided face validity does not meet current NQF requirements, however, face validity is not needed when empirical validity testing is provided.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

Preliminary rating for reliability:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
Preliminary rating for validity:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- na
- no concerns about reliability. no risk adjustment.
- Specifications are clear, as is the calculation algorithm. Consistent implementation is highly likely.
- Data elements are clear. No concerns.
- reliability - specifications are appropriate
- I have no specific concerns for reliability.
- The description of the calculation algorithm seems clear.
- No concerns with the reliability of the measure specifications. No concerns that the measure can be consistently implemented.
- No concerns related to reliability

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- developer used a beta-binominal model to assess the signal-to-noise ratio
- none.
- Reliability testing showed high reliability for both commercial and medicaid plans. No concerns regarding reliability
- No concerns.
- no
- It appears that the measure stands up well in all comparisons.
- Reliability was analyzed by the beta binomial method (signal to noise reliability). Results using this method are consistently better than 0.7, the minimal threshold for reliability.
- No concerns about the reliability of the measure.
- No concerns

2b1. Validity -Testing: Do you have any concerns with the testing results?

- conducted correlation analysis with three other measures

- no
- None
- No concerns
- no
- One question that I'd like for developer to discuss was the relationship with Low Back Pain Imaging studies. I see the correlation but wonder if the developer can explain what might be going on?
- Construct validity was tested using Pearson correlations on HEDIS measures - avoidance of antibiotic treatment for acute bronchitis/bronchiolitis, use of imaging studies for low back pain, and antibiotic utilization. There are no significant concerns with these results.
- No concerns with the validity testing conducted by the measure developer.
- No concerns

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- Exclusion of comorbid conditions?
- exclusions are appropriate. no risk adjustment
- n/a
- Exclusions are consistent with evidence.
- no major concerns
- At present, I see no issues with exclusions or risk adjustment.
- No risk adjustment or stratification
- No threats to validity. Exclusions are appropriate.
- No concerns

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- na
- testing shows good reliability.
- 2b4 - Yes; 2b5 - Yes; 2b6 - no
- No threats to validity.
- no
- No issues with threats to validity. It appears that this measure can be implemented to be valid.
- Meaningful Differences - Tested by inter-quartile range tests that were statistically significant. Comparability of performance scores - There is only one set of specifications. Missing data - NCQA auditors did not find sources of missing data for this measure in evaluating the last of the three years of data submitted in the maintenance submission. Exclusions were reasonable and limited.

- Measure results should result in the identification of any meaningful differences. Results should be comparable. No concerns with missing data in reporting the measure.
- No concerns

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- Data is collected during the provision of care and coded by someone other than person obtaining original information
- All data elements are in defined fields in electronic claims.
- The developer conducts audits for all HEDIS collection and reporting processes. An independent audit of HEDIS process to verify integrity of HEDIS collection and reporting system is conducted and a Policy Clarification Support System is used to generate ongoing feedback from measure users.
- The developer notes that noncommercial uses do not require the consent of the measure developer. However, commercial use of the measure requires the prior written consent of NCQA.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

1. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- data elements are in defined fields in electronic claims
- All data elements generated during care. no data collection concerns.
- No concerns about data collection, as it is electronic data. Diagnosis accuracy may be less accurate depending on the coding accuracy at the point of care by providers. Not overly concerned.
- Data elements are routinely generated during care delivery. No concerns
- no major concerns
- Continued measure seems to be feasible.
- The data elements are routinely generated and available from electronic forms. The process is subject to the HEDIS Compliance Audit. The data collection process for this measure has been operating for three years.
- All data for the measure are in electronic claim fields. All are routinely generated during care delivery. The data collection strategy is operational.
- Data elements are routinely generated in the delivery of care

Criterion 4: Usability and Use

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

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- The measure is used in the following programs: NCQA Health Plan Rating/Report Cards; NCQA State Of Health Care Quality; Qualified Health Plan (QHP) Quality Rating System (QRS); CDC Measuring Outpatient Antibiotic Prescribing; Quality Payment Program; NCQA Health Plan Accreditation; NCQA Quality Compass

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others [vetting]

a. NCQA publishes HEDIS results annually and presents at various conferences and webinars.
b. Technical assistance is provided on measures through the developer's Policy Clarification Support System.
c. NCQA utilizes a consensus-based process to obtain broad input on the measure from several multi-stakeholder advisory panels, public comment posting, and questions submitted to the Policy Clarification Support System.
d. The developer notes that they have not received any feedback which would indicate any barriers to measure implementation.

Questions for the Committee:

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

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

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer notes that improvement results cannot be reported for this measure due to the change in measure denominator age ranges from 2018 to 2019.
- The developer notes that for 2019, the average performance rate was 78.6% for commercial plans and 86.9% for Medicaid plans.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer noted that the intended benefits of this measure outweigh the potential harms.
- The developer does not list any potential harms.

Potential harms

- The developer states there were no identified potential harms for this measure during testing or since implementation.

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Publicly reported and currently used in accountability programs
- data is used in multiple accountability programs. feedback is obtained via multiple mechanisms.
- Measure is publicly reported by NCQA, Qualified Health Plan Quality Rating System and CDC. Feedback is consensus based on stakeholder advisory groups and public comment. Developer states they have not received feedback that the measure is not able to be successfully implemented.
- Publicly reported and used in accountability programs listed in application
- no major concerns
- Meets the requirements for Use.
- The measure has been reported through NCQA and CDC publications. it is used for NCQA health plans accreditation and quality improvement projects. There are ongoing processes to solicit and evaluate feedback on this measure from stakeholders and the public.

- NCQA uses a consensus-based process to obtain broad input on the measure from stake-holder advisory panels, public comment postings and questions submitted to the PCS system. No concerns with the use-feedback process.
- Currently used in multiple public reporting and accountability programs

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- improvement not reported for this measure due to the change in measure denominator
- change in measurement definitions from 2018 to 2019 limit ability to show improvement . no unintended consequences. benefits outweigh harms.
- The intent to drive improvement is in part by payment adjustments to promote higher quality care. Withholding antibiotics in situations where they are indicated is a potential unintended consequence but with low likelihood. Benefits of far fewer antibiotics outweigh the risks of not treating at the population level.
- No harms
- no concerns
- Appears to meet the requirements for Usability.
- Establishing performance improvement is complicated by the change in patient ages included and the change from member based measure to an episode based measure. There are no identified unintended consequences with this measure.
- Due to measure specification changes including age range changes from 2018 to 2019 the rates are not currently comparable to prior years. The results are reported in Quality Compass, used for accreditation, and used to compare programs and health plan results. Benefits outweigh the harms. The measure has been vetted in real-life settings. The results can be used to improve healthcare quality.
- No concerns related to unintended consequences

Criterion 5: Related and Competing Measures

Related or competing measures

- 0058: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (AAB)

Harmonization

- The developer indicated that this measure has been harmonized with NQF 0058.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- bronchitis measure has been harmonized.
- Harmonized with NQF 0058
- Harmonized with competing measure NQF0058. Expanded to include adults.
- appropriately harmonized with 0058
- Harmonized with other related measures.

- 0058 Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis is listed. There are parallels in methodology and goals but 0069 focuses on upper respiratory infections and 0058 is focused on types of lower respiratory infections. The developer noted that 0058 has been harmonized with this measure (but no details or explanation were provided).
- Competing measure #0058: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (AAB). The measures have been harmonized.
- Related measures have been harmonized

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/21/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0069

Measure Title: Appropriate Treatment for Upper Respiratory Infection

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed** (NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_0069" document, items S.1-S.22

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**
   - The population of this measure has been expanded to beneficiaries aged 3 months or older. The measure previously included beneficiaries three months to 18 years old.
   - The measure was also changed to an episode based measure.

- No concerns.

RELIABILITY: TESTING

Submission document: "MIF_0069" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   - The developer used a beta-binominal model to assess the signal-to-noise ratio of the measure score.
   - The reliability of the measure score was assessed using 2019 HEDIS data.
   - A minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   - The developer provided mean signal-to-noise reliability for commercial and Medicaid plans for all ages as well as ages 3 months-17 years; age 18-64; and, 65 years and older.

Appropriate Treatment for Upper Respiratory Infection	Point estimate: Mean Signal-To-Noise Reliability (Commercial)	Point estimate: Mean Signal-To-Noise Reliability (Medicaid)
Appropriate Treatment for Upper Respiratory Infection (Total)	0.983	0.992
Appropriate Treatment for Upper Respiratory Infection (age 3 Months-17 Years)	0.955	0.993
Appropriate Treatment for Upper Respiratory Infection (age 18-64)	0.980	0.979
Appropriate Treatment for Upper Respiratory Infection (age 65+)	0.900	0.945

   - The developer states that the reliability estimates provided indicate good reliability.
8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

- A suitable method was used for reliability testing and scores of 0.992 for Medicaid and 0.983 for commercial plans indicate good reliability.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

Submission document: Testing attachment, section 2b2.

- This measure excludes: episodes with a history of select comorbid conditions, patients with a history of antibiotic use, or presence of a competing diagnosis.
- The developer noted that exclusions are tested during measure development rather than reporting.
- The developer reported the following percentages for excluded populations:
    - Comorbid conditions:
        - Commercial: 1.1%
        - Medicaid: 0.8%-2.7%
        - Medicare: 0.8%-2.7%
    - Competing diagnosis:
        - Commercial: 25.7%
        - Medicaid: 18%-25.6%
        - Medicare: 12.3%-14.5%
    - History of antibiotic use:
        - Commercial: 8.4%
        - Medicaid: 8.3%-11.1%
        - Medicare: 12.3%-15.7%
- The developer concluded that removing these exclusion from the denominator was appropriate given the magnitude of the exclusions.
- No concerns.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

- To demonstrate meaningful differences in performance, the developer calculated an inter-quartile range for both commercial and Medicaid health plans.

- For commercial plans, the IQR was 11%, which represents an average of 494 more patients receiving appropriate treatment for upper respiratory infections in high-performing plans compared to low-performing plans.
- For Medicaid plans, the IQR was 6%, which represents an average of 1,140 more patients receiving appropriate treatment for upper respiratory infections in high-performing plans compared to low-performing plans.
- No concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    - Not Applicable
15. **Please describe any concerns you have regarding missing data.**
    Submission document: Testing attachment, section 2b6.
    - The developer indicated that it audits the diagnostic and procedure code fields for this measure.
    - The developer reported that no missing data was found during the audit.
    - No concerns.
16. **Risk Adjustment**

    16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

    16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**
    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

    16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

    16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

    16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

    16d. **Risk adjustment summary:**

    16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

    16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

    16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

    16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

    16d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

    16e**. Assess the risk-adjustment approach**
    - Not Applicable

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**
    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**
18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
20. **Method of establishing validity of the measure score:**

☐ **Face validity**

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

☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

- The developer assessed construct validity on 2019 HEDIS data by calculating Pearson Correlation Coefficient between this measure and the HEDIS measure Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis and Use of Imaging Studies for Low Back Pain; hypothesizing a positive correlation.
- The developer also compared this measure to the HEDIS measure Antibiotic Utilization; hypothesizing a negative correlation.
- The developer stated that face validity was provided but it did not meet NQF requirements. However, since empirical validity testing was provided, face validity is not needed.

22. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

- The developer found a positive correlation between this measure and Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis.
    - Medicaid: Correlation coefficient = 0.68, $p < 0.001$
    - Commercial: Correlation coefficient = 0.68, $p < 0.001$
- The developer found a positive correlation between this measure and Use of Imaging Studies for Low Back Pain.
    - Medicaid: Correlation coefficient = 0.41, $p < 0.001$
    - Commercial: Correlation coefficient = 0.622, $p < 0.001$
- The developer found a negative correlation between this measure and Antibiotic Utilization.
    - Medicaid: Correlation coefficient = -0.73, $p < 0.001$
    - Commercial: Correlation coefficient = -0.74, $p < 0.001$
- Face Validity: Results from multiple multi-stakeholder measurement advisory panels, as well as those submitting to public comment, indicate that the measure as specified has sufficient face validity and will accurately differentiate quality across providers.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

NOTE that data element validation from the literature is acceptable.

Submission document*: Testing attachment, section 2b1.*

☐ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

- Score level testing was conducted.
- Correlation analysis demonstrated construct validity of this measure.
- The developer provided data related to exclusions and meaningful differences. The developer noted that they did not find any issues with missing data.
- No concerns

27. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

    Not Applicable

ADDITIONAL RECOMMENDATIONS

28. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    No concerns

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

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

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

nqf_evidence_attachment_7.1_URI_10.23.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a. Evidence (subcriterion 1a)

1a.1. This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: Appropriate Treatment for Upper Respiratory Infection – this measure assesses the percentage of episodes with a diagnosis of upper respiratory infection that did not result in an antibiotic dispensing event.

☒ Appropriate use measure: Appropriate use of antibiotics

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

2020 Submission

Reduction in the inappropriate dispensing of antibiotics for upper respiratory infections, often referred to as "a common cold" >> fewer strains of antibiotic-resistant bacteria >>reduction in community-acquired antibiotic resistant infections

The intended result of efforts to decrease indiscriminant antibiotic use in the ambulatory setting is to reduce (and preferably reverse) the increase in antibiotic-resistant bacteria, widely considered to be a threat to public health. The Centers for Disease Control and Prevention underscores the importance of decreasing community

use of antibiotics as an important strategy for combating the increase in community-acquired antibiotic resistant infections.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

2020 Submission

This measure assesses whether patients 3 months of age and older with a diagnosis of upper respiratory infection (URI) were not dispensed an antibiotic prescription. The measure intent is to decrease unnecessary prescribing of antibiotic treatment. Too often, antibiotics are prescribed inappropriately, which can lead to antibiotic resistance (when antibiotics can no longer cure bacterial infections).

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	2020 Submission Spurling, G. K., Del Mar, C. B., Dooley, L., Foxlee, R., & Farley, R. (2017). Delayed antibiotic prescriptions for respiratory infections. *The Cochrane database of systematic reviews*, *9*(9), CD004417. https://doi.org/10.1002/14651858.CD004417.pub5 2013 Submission Principles of Appropriate Antibiotic Use: Part II. Nonspecific Upper Respiratory Tract Infections. Am Fam Physician. 2001 Sep 15;64(6):1098-1099.

Systematic Review	Evidence
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	2020 Submission This systematic review concluded that for many clinical outcomes associated with respiratory infections, there were no differences between three prescribing strategies (i.e. immediate antibiotics, delayed antibiotics, no antibiotics). Not prescribing antibiotics at all was shown to reduce antibiotic use compared to delaying antibiotic prescription or immediate antibiotic prescription. Delaying antibiotics for people with acute respiratory infection was not shown to be different to no antibiotics in terms of symptom control and disease complications. 2013 Submission These recommendations apply only to immunocompetent adults with no important comorbid conditions, such as pulmonary or cardiac disease - In previously healthy adults, the diagnosis of nonspecific upper respiratory tract infection should be used to denote an acute infection that is typically viral in origin and in which sinus, pharyngeal and lower airway symptoms, although frequently present, are not prominent. Most cases of uncomplicated upper respiratory tract infection in adults resolve spontaneously. Symptoms typically last one to two weeks, and most patients feel better within the first week. These infections are predominantly viral in origin, and complications, such as bacterial rhinosinusitis or bacterial pneumonia, are rare. - Antibiotic treatment of adults with nonspecific upper respiratory tract infection is not recommended because it does not enhance illness resolution or alter the rates of uncommon complications. - Purulent nasal discharge and sputum do not predict bacterial infection and patients with these symptoms do not benefit from antibiotic treatment. Antibiotic therapy does not decrease the duration of symptoms or lost work time, or prevent complications.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	2020 Submission Overall, the quality of the evidence for the systematic review was **moderate** according to GRADE assessment. **Moderate quality:** We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. 2013 Submission NA

Systematic Review	Evidence
Provide all other grades and definitions from the evidence grading system	2020 Submission **GRADE Working Group grades of evidence** **High quality:** We are very confident that the true effect lies close to that of the estimate of the effect. **Low quality:** Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect. **Very low quality:** We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect. 2013 Submission Guideline was not graded, however, the Centers for Disease Control and Prevention (CDC) deemed the inappropriate use of antibiotics an important public health issue that warranted the development of this guideline. To address this issue, CDC assembled a panel of national health experts, including physicians with expertise in internal, family, emergency and infectious diseases medicine, to develop evidence based guidelines for evaluating and treating adults with acute respiratory illness. The goal of the guidelines put together by the CDC and other members of the panel is to provide physicians with practical strategies for limiting antibiotic use to patients who are most likely to benefit. In addition to the CDC, the principles outlined in the above mentioned guidelines have been endorsed by the American Academy of Family Physicians (AAFP), the American College of Physicians–American Society of Internal Medicine (ACP-ASIM), and the Infectious Diseases Society of America (IDSA)
Grade assigned to the **recommendation** with definition of the grade	2020 Submission NA 2013 Submission NA
Provide all other grades and definitions from the recommendation grading system	2020 Submission NA 2013 Submission NA

Systematic Review	Evidence
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	2020 Submission This is a living systematic review which is continually updated. For this 2017 update, the authors added one new trial involving 405 participants with uncomplicated acute respiratory infection. Overall, this review included 11 studies with a total of 3,555 participants. These 11 studies involved acute respiratory infections including acute otitis media (three studies), streptococcal pharyngitis (three studies), cough (two studies), sore throat (one study), common cold (one study), and a variety of respiratory tract infections, or RTIs (one study). Five studies focused on children, two focused on adults, and four included both adults and children. Six studies were conducted in a primary care setting, three in pediatric clinics, and two in emergency departments. All studies were randomized controlled trials, involving participants of all ages having an RTI, and comparing two or three antibiotic prescribing strategies. Studies were well reported and appeared to be of moderate quality. 2013 Submission NA
Estimates of benefit and consistency across studies	2020 Submission Studies have consistently shown that antibiotics are not appropriate for treating upper respiratory infections, many of which are caused by a virus. Overprescribing can lead to antibacterial resistance. This 2017 SR update assesses studies comparing three different antibiotic prescribing strategies for RTIs, with the primary finding being there is no difference among them (e.g. no benefit to immediate or delayed antibiotic prescribing compared to no prescribing). 2013 Submission 1c.25 Quantity: Moderate 1c.26 Quality: Moderate1c.27 Consistency: Moderate
What harms were identified?	2020 Submission In a small number of studies, evidence suggests some benefit of prescribing antibiotics over not prescribing them. For example, in one instance, delayed antibiotic prescribing led to a small reduction in the duration of cold symptoms, compared to no antibiotics. 2013 Submission NA

Systematic Review	Evidence
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	2020 Submission DeGeorge, K. C., Ring, D. J., & Dalrymple, S. N. (2019). Treatment of the Common Cold. *American family physician*, *100*(5), 281–289. https://www.aafp.org/afp/2019/0901/p281.html#afp20190901p281-b46 Literature published since this 2017 systematic review is consistent with the recommendation to delay or avoid antibiotics for upper respiratory infections. 2013 Submission NA

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	2020 Submission Kenealy, T., & Arroll, B. (2013). Antibiotics for the common cold and acute purulent rhinitis. *The Cochrane database of systematic reviews*, *2013*(6), CD000247. https://doi.org/10.1002/14651858.CD000247.pub3
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	2020 Submission There is no evidence of benefit from antibiotics for the common cold or for persisting acute purulent rhinitis in children or adults. There is evidence that antibiotics cause significant adverse effects in adults when given for the common cold and in all ages when given for acute purulent rhinitis. Routine use of antibiotics for these conditions is not recommended.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	2020 Submission No grade assigned to the evidence
Provide all other grades and definitions from the evidence grading system	2020 Submission The systematic review did not include an evidence grading system
Grade assigned to the **recommendation** with definition of the grade	2020 Submission NA
Provide all other grades and definitions from the recommendation grading system	2020 Submission NA

Systematic Review	Evidence
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	2020 Submission This updated review included 11 studies. Six studies contributed to one or more analyses related to the common cold, with up to 1047 participants. Five studies contributed to one or more analyses relating to purulent rhinitis, with up to 791 participants. One study contributed only to data on adverse events and one met the inclusion criteria but reported only summary statistics without providing any numerical data that could be included in the meta-analyses. Interpretation of the combined data is limited because some studies included only children, or only adults, or only males; a wide range of antibiotics were used and outcomes were measured in different ways. There was a moderate risk of bias because of unreported methods details or because an unknown number of participants were likely to have chest or sinus infections. All trials were standard parallel design randomized controlled trials.
Estimates of benefit and consistency across studies	2020 Submission These randomized controlled trials comparing any antibiotic therapy against placebo in people with symptoms of acute upper respiratory tract infection or acute purulent rhinitis consistently showed no benefit from antibiotics.
What harms were identified?	2020 Submission Authors conclude that antibiotics can cause side effects such as diarrhea. Most notably, unnecessary antibiotic use for upper respiratory infections leads to bacteria becoming resistant to antibiotics.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	2020 Submission Conclusions from studies published since this systematic review (summarized in table above) have not changed.

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	Short S, Bashir H, Marshall P, Miller N, Olmschenk D, Prigge K, Solyntjes L. Institute for Clinical Systems Improvement. Diagnosis and Treatment of Respiratory Illness in Children and Adults. Updated September 2017. https://www.icsi.org/wp-content/uploads/2019/01/RespIllness.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	The ICSI work group does not recommend antibiotics for treatment of common cold symptoms in children and adults

Systematic Review	Evidence
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**Low:** The true effect might be markedly different from the estimated effect. (ICSI utilizes the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology system.)
Provide all other grades and definitions from the evidence grading system	**High:** The authors have a lot of confidence that the true effect is similar to the estimated effect. **Moderate:** The authors believe that the true effect is probably close to the estimated effect. **Very low:** The true effect is probably markedly different from the estimated effect.
Grade assigned to the **recommendation** with definition of the grade	**Strong:** Confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects
Provide all other grades and definitions from the recommendation grading system	**Weak:** The desirable effects of adherence to a recommendation probably outweigh the desirable effects (less confident)
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	This recommendation is based on an updated Cochrane systematic review, which included randomized controlled trials with 1,047 patients. (Kenealy & Arroll, 2013; cited above)
Estimates of benefit and consistency across studies	This is a common conclusion of studies of this type. Not treating with antibiotics eliminates the possible side effects of antibiotics such as nausea, vomiting, allergic reactions, and *Clostridium Difficile* infection. In addition, better stewardship of antibiotics helps reduce potential for antibiotic resistance.
What harms were identified?	No harms were identified.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Conclusions from studies published since this systematic review (summarized in table above) have not changed.

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

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

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

This measure assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection (URI) were not dispensed an antibiotic prescription. The improvement in quality envisioned by the use of this measure is to decrease unnecessary prescribing of antibiotic treatment for upper respiratory infection. Too often, antibiotics are prescribed inappropriately, which can lead to antibiotic resistance (when antibiotics can no longer cure bacterial infections).

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. It includes number of health plans, percentiles, mean, min, max and standard deviations.

Data are summarized at the health plan level (i.e. "N" represents the number of health plans)

The rate is reported as an inverted rate (i.e. 1- numerator/denominator) to reflect the number of people in the health plans that were not dispensed an antibiotic. Data are stratified by year and product line (i.e. commercial, Medicare, Medicaid)

Commercial

2020 TOTAL RATE

N 404

Mean 78.66

Std Dev 8.59

Min 48.51

P10 67.59

P25 73.48

P50 80.11

P75 84.82

P90 89.00

Max 96.17

IQR:11.34

Medicaid

2020 TOTAL RATE

N 222

Mean 86.92

Std Dev 6.79

Min 49.96

P10 78.79

P25 85.25

P50 87.78

P75 91.09
P90 93.53
Max 100.00
IQR: 5.84
Commercial
2019 TOTAL RATE
N 378
Mean 90.06
Std Dev 5.97
Min 58.42
P10 82.29
P25 87.89
P50 91.48
P75 93.72
P90 95.84
Max 100.00
IQR: 5.83
Medicaid
2019 TOTAL RATE
N 199
Mean 90.45
Std Dev 6.36
Min 65.15
P10 82.04
P25 87.91
P50 91.85
P75 94.88
P90 96.79
Max 100.00
IQR:6.97
Commercial
2018 TOTAL RATE
N 384
Mean 88.26
Std Dev 7.09
Min 44.11
P10 79.21
P25 85.47
P50 90.18

P75 92.50

P90 94.93

Max 100.00

IQR: 7.03

Medicaid

2018 TOTAL RATE

N 207

Mean 89.08

Std Dev 7.11

Min 60.25

P10 80.55

P25 86.63

P50 90.42

P75 93.77

P90 95.94

Max 100.00

IQR: 7.14

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

N/A. Performance data are included in 1b.2

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities if the data are available to a plan. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. Our work is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.1,2 This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

HEDIS includes two measures that can be used as tools for assessing race/ethnicity and language needs of a plan's population: Race/Ethnicity Diversity of Membership and the Language Diversity of Membership. These measures promote standardized methods for collecting these data and follow Office of Management and Budget and National Academy of Medicine guidance for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf

2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-based-purchasing-programs-reports

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

Data on racial and ethnic disparities in antibiotic prescribing for upper respiratory infection are limited. However, a study by Goyal et al. (2017) found that non-Latino white children seeking treatment for viral infections in the emergency department setting were approximately twice as likely to be prescribed an unnecessary antibiotic compared with their black or Latino counterparts. The authors recommend that future research explore some of the reasons that could explain these differences in antibiotic prescribing, such as parental pressure and implicit bias.

Goyal MK et al. Racial and ethnic differences in antibiotic use for viral illness in emergency departments. Pediatrics 2017 Aug 2; [e-pub]. (http://dx.doi.org/10.1542/peds.2017-0203)

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, **as specified**, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

De.5. Subject/Topic Area *(check all the areas that apply):*

Infectious Diseases (ID), Infectious Diseases (ID) : Pneumonia and respiratory infections

De.6. Non-Condition Specific*(check all the areas that apply):*

Safety : Overuse

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Children

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

N/A

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

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment**:** 0069_URI_Fall_2020_Value_Sets.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

The measure was recently expanded to include adults (previously limited to children 3 months to 18 years of age). The measure as currently specified assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection (URI) were not dispensed an antibiotic prescription. NCQA expanded the age range for this measure based on stakeholder and measurement advisory panel feedback that unnecessary prescribing affects the general population and therefore the measure should include all ages and product lines. The measure includes patients enrolled in commercial, Medicaid, and Medicare health plans.

The measure was also changed from a member-based measure to an episode-based measure. The member-based denominator resulted in members with multiple URIs during the measurement period counting only once. An episode-based measure captures more episodes of potentially inappropriate antibiotic prescribing.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The numerator of the measure includes the number of dispensed prescriptions for an antibiotic medication on or 3 days after the Episode Date.

S.5. 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, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Dispensed antibiotic medications (Table CWP Antibiotic Medications) on or within 3 days after an outpatient, telephone, e-visit or virtual check-in, an observation visit or ED encounter for upper respiratory infection (URI) during the intake period. The measure is reported as an inverted rate (1-numerator/denominator); a higher rate is better.

CWP Antibiotic Medications

Aminopenicillins: Amoxicillin, Ampicillin

Beta-lactamase inhibitors: Amoxicillin-clavulanate

First generation cephalosporins: Cefadroxil, Cefazolin, Cephalexin

Folate antagonist: Trimethoprim

Lincomycin derivatives: Clindamycin

Macrolides: Azithromycin, Clarithromycin, Erythromycin, Erythromycin ethylsuccinate, Erythromycin lactobionate, Erythromycin stearate

Natural penicillins: Penicillin G potassium, Penicillin G benzathine, Penicillin G sodium, Penicillin V potassium

Penicillinase-resistant penicillins: Dicloxacillin

Quinolones: Ciprofloxacin, Levofloxacin, Moxifloxacin, Ofloxacin

Second generation cephalosporins: Cefaclor, Cefprozil, Cefuroxime

Sulfonamides: Sulfamethoxazole-trimethoprim

Tetracyclines: Doxycycline, Minocycline, Tetracycline

Third generation cephalosporins: Cefdinir, Cefixime, Cefpodoxime, Ceftibuten, Cefditoren, Ceftriaxone

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Episodes for members 3 months of age and older as of July 1 of the year prior to the measurement year who had an outpatient, telephone, e-visit or virtual check-in, an observation visit or ED encounter with a diagnosis of upper respiratory infection (URI) during the intake period (July 1st of the year prior to the measurement year to June 30th of the measurement year).

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Follow the steps below to identify the eligible population:

Members who had an outpatient visit (Outpatient Value Set), a telephone visit (Telephone Visits Value Set), an e-visit or virtual check-in (Online Assessments Value Set) an observation visit (Observation Value Set) or an ED visit (ED Value Set) during the Intake Period, with a diagnosis of URI (URI Value Set).

The member must be continuously enrolled without a gap in coverage from 30 days prior to the Episode Date through 3 days after the Episode Date (34 total days).

Deduplicate eligible episodes. If a member has more than one eligible episode in a 31-day period, include only the first eligible episode. For example, if a member has an eligible episode on January 1, include the January 1 visit and do not include eligible episodes that occur on or between January 2 and January 31; then, if applicable, include the next eligible episode that occurs on or after February 1. Identify visits chronologically, including only one per 31-day period.

CWP-C: Antibiotic Medications

Aminopenicillins: Amoxicillin, Ampicillin

Beta-lactamase inhibitors: Amoxicillin-clavulanate

First generation cephalosporins: Cefadroxil, Cefazolin, Cephalexin

Folate antagonist: Trimethoprim

Lincomycin derivatives: Clindamycin

Macrolides: Azithromycin, Clarithromycin, Erythromycin, Erythromycin ethylsuccinate, Erythromycin lactobionate, Erythromycin stearate

Natural penicillins: Penicillin G benzathine, Penicillin G potassium, Penicillin G sodium, Penicillin V potassium

Penicillinase-resistant penicillins: Dicloxacillin

Quinolones: Ciprofloxacin, Levofloxacin, Moxifloxacin, Ofloxacin

Second generation cephalosporins: Cefaclor, Cefprozil, Cefuroxime

Sulfonamides: Sulfamethoxazole-trimethoprim

Tetracyclines: Doxycycline, Minocycline, Tetracycline

Third generation cephalosporins: Cefdinir, Cefixime, Cefpodoxime, Ceftibuten, Cefditoren, Ceftriaxone

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Exclude visits that result in an inpatient stay.

Exclude Episode Dates when the member had a claim/encounter with any diagnosis for a comorbid condition during the 12 months prior to or on the Episode Date.

Exclude Episode Dates where a new or refill prescription for an antibiotic medication was filled 30 days prior to the Episode Date or was active on the Episode Date.

Exclude Episode Dates where the patient had a claim/encounter with a competing diagnosis on or three days after the Episode Date.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

Exclude visits that results in an inpatient stay (Inpatient Stay Value Set)

Exclude Episode Dates when the member had a claim/encounter with any diagnosis for a comorbid condition during the 12 months prior to or on the Episode Date. A code from any of the following meets criteria for a comorbid condition:

- HIV Value Set.
- HIV Type 2 Value Set.
- Malignant Neoplasms Value Set.
- Other Malignant Neoplasm of Skin Value Set
- Emphysema Value Set.
- COPD Value Set.
- Comorbid Conditions Value Set.
- Disorders of the Immune System Value Set

Exclude for Negative Medication History: No pharmacy claims for either new or refill prescriptions for an antibiotic drug listed below in the 30 days prior to Episode Date, or was active on Episode Data :

CWP-C: Antibiotic Medications

Aminopenicillins: Amoxicillin, Ampicillin

Beta-lactamase inhibitors: Amoxicillin-clavulanate

First generation cephalosporins: Cefadroxil, Cefazolin, Cephalexin

Folate antagonist: Trimethoprim

Lincomycin derivatives: Clindamycin

Macrolides: Azithromycin, Clarithromycin, Erythromycin, Erythromycin ethylsuccinate, Erythromycin lactobionate, Erythromycin stearate

Natural penicillins: Penicillin G benzathine, Penicillin G potassium, Penicillin G sodium, Penicillin V potassium

Penicillinase-resistant penicillins: Dicloxacillin

Quinolones: Ciprofloxacin, Levofloxacin, Moxifloxacin, Ofloxacin

Second generation cephalosporins: Cefaclor, Cefprozil, Cefuroxime

Sulfonamides: Sulfamethoxazole-trimethoprim

Tetracyclines: Doxycycline, Minocycline, Tetracycline

Third generation cephalosporins: Cefdinir, Cefixime, Cefpodoxime, Ceftibuten, Cefditoren, Ceftriaxone

Exclude Episodes where there is a claim/encounter for a competing diagnosis on or 3 days after the Episode Date. A code from either of the following meets criteria for a competing diagnosis:

- Pharyngitis Value Set.
- Competing Diagnosis Value Set.

(See corresponding Excel document for the value sets referenced above)

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Measure is stratified by age:

3 months – 17 years

18 - 64 years

65 years and older

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Other

If other: The measure is reported as an inverted rate [1 – (numerator/denominator)], therefore a higher score represents the proportion of patients for whom antibiotics were not prescribed.

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or 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; time period for data, aggregating data; risk adjustment; etc.)*

Episode Date is defined as the date of service for any outpatient, telephone, observation or ED visit, e-visit or virtual check-in during the Intake Period with a diagnosis of URI.

Step 1 Determine the eligible population. To do so, identify all patients who had an outpatient, telephone, e-visit or virtual check-in or ED visit with a diagnosis of URI during the Intake Period.

Step 2 Determine all URI Episode Dates during the intake period. For each patient identified in step 1, determine all outpatient, telephone, observation or ED claims/encounters or e-visits and virtual check-ins with a URI diagnosis.

Step 3 Test for Negative Comorbid Condition History. Exclude Episode Dates when the patient had a claim/encounter with any diagnosis for a comorbid condition during the 12 months prior to or on the Episode Date.

Step 4 Test for Negative Medication History. Exclude Episode Dates where a new or refill prescription for an antibiotic medication was filled 30 days prior to the Episode Date or was active on the Episode Date.

Step 5 Test for Negative Competing Diagnosis. Exclude Episode Dates where the patient had a claim/encounter with a competing diagnosis on or three days after the Episode Date.

Step 6 Calculate continuous enrollment. The patient must be continuously enrolled without a gap in coverage from 30 days prior to the Episode Date through 3 days after the Episode Date (34 total days).

Step 7 Deduplicate eligible episodes. If a patient has more than one eligible episode on a 31-day period, include only the first eligible episode. (provides denominator)

Step 8 Calculate numerator - number of dispensed prescriptions for an antibiotic medication from the Antibiotic Medication list on or 3 days after the episode date

Step 9 Calculate rate numerator/denominator

Step 10 Subtract the rate calculated in Step 9 from 1 to invert the measure result to represent appropriate treatment for upper respiratory infection (i.e., antibiotic not prescribed). The measure is reported as an inverted rate (i.e., 1 – numerator/denominator) to reflect the number of episodes not associated with a dispensed antibiotic (higher is better).

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via NCQA's online data submission system.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Health Plan

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Emergency Department and Services, Outpatient Services

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

URI_0069_Testing_Form_-637400921815073774.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0069
Measure Title: Appropriate Treatment for Upper Respiratory Infection
Date of Submission: 8/3/2020

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

D1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

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

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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

2020 Submission

This measure was tested using administrative claims data from Medicaid and commercial plans nationwide that reported data for the annual Healthcare Effectiveness Data and Information Set (HEDIS)

2013 Submission

N/A

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

2020 Submission

January 1 – December 31, 2019.

2013 Submission

2009, 2010, 2011

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other:	☐ other:

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

2020 Submission

This measure was recently expanded to include members 3 months of age and older (previously limited to children 3 months- 18 years of age). The measure as currently specified assesses whether members 3 months of age and older with a diagnosis of upper respiratory infection (URI) were not dispensed an antibiotic prescription. The measure includes patients enrolled in commercial, Medicaid, and Medicare health plans. The

intent of the measure is to decrease unnecessary prescribing of antibiotic treatment. Too often, antibiotics are prescribed inappropriately, which can lead to antibiotic resistance (when antibiotics can no longer cure bacterial infections).
This measure is reported as an inverted rate [1–(numerator/eligible population)]. A higher rate indicates appropriate URI treatment (i.e., the proportion of episodes that did *not* result in an antibiotic dispensing event).
Sample for measure score reliability testing and construct validity testing: The measure score reliability was calculated from HEDIS data that included 404 commercial health plans and 222 Medicaid health plans. The sample included all commercial, and Medicaid plans submitting data to NCQA for this HEDIS measure. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity

The measure was assessed for face validity through two independent panels of experts:

- NCQA's Committee on Performance Measurement (CPM) oversees measures used in NCQA programs and includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is composed of 17 independent members that reflect the diversity of constituencies that performance measurement serves. The CPM's recommendations are reviewed and approved by NCQA's Board of Directors.
- NCQA's Antibiotic Overuse Measurement Advisory Panel is composed of 8 independent members representing hospitals, public policy research, public health and universities. This panel oversees HEDIS antibiotic use measures to align with current evidence-based guidelines and practices.

2013 Submission

The data exist in HEDIS Performance Measurement data for 2011.
Number of commercial health plans, 2011= 405
Number of Medicaid health plans, 2011= 154
It was a geographically diverse sample of health plans.

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

2020 Submission

Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicaid). Below is a description of the sample. It includes number of health plans included in HEDIS data collection and the median eligible population for the measure across health plans. The eligible population for this measure is based on episodes. NCQA does not collect patient-level data from health plans but can provide age/sex information.

Table 1. Median eligible population for the *Appropriate Treatment for Upper Respiratory Infection* measure by plan type, calendar year 2019 data

Product Type	Number of Plans	Median number of eligible episodes per plan
Commercial	404	4,493
Medicaid	222	19,004

2013 Submission

The data exist in HEDIS Performance Measurement data for 2011.

Number of commercial health plans, 2011= 405

Number of Medicaid health plans, 2011= 154

It was a geographically diverse sample of health plans.

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

2020 Submission

There were no differences in the data used for reliability, construct validity or meaningful differences in performance testing. As described above in Section 1.5, two multi-stakeholder expert panels assessed face validity.

2013 Submission

NA

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

2020 Submission

We did not assess data by social risk factors. Social risk factor data were not available in reported results. This measure is specified for members aged 3 months and older for all product lines. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. This is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.1,2This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

2013 Submission

NA

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

2020 Submission

We utilized the methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) to calculate signal-to-noise reliability. This methodology uses the Beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences across reporting entities (plans, physicians, etc.) in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the Appropriate Treatment for Respiratory Infection (URI) measure. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities.

For the URI measure, Medicaid and commercial health plans are the reporting entity. For the formulas and explanations below, we use the health plan as the reporting entity.

The formula for signal-to-noise reliability is:

Signal-to-noise reliability = $\sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$

Therefore, we need to estimate two variances: 1) variance between plans ($\sigma^2_{\text{plan-to-plan}}$); 2) variance within plans (σ^2_{error}).

1. Variance between plans = $\sigma^2_{\text{plan-to-plan}} = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

   α and β are two shape parameters of the Beta-Binomial distribution, $\alpha > 0$, $\beta > 0$
2. Variance within plans: $\sigma^2_{\text{error}} = \hat{p}(1- \hat{p})/n$

   $\hat{p}$ = observed rate for the plan

   n = plan-specific denominator for the observed rate (in this case, the number of eligible episodes of upper respiratory infection)

Using Adams' 2009 methodology, we estimated the reliability for each reporting entity, then averaged these reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability. We label this point estimate "mean signal-to-noise reliability". The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.

Along with the point estimate of mean signal-to-noise reliability, we are also providing:

1. The standard error (SE) and 95% confidence interval (95% CI) of the mean signal-to-noise reliability for all plans and stratified by the denominator size (number of eligible members per plan). The SE and 95% CI of the mean signal-to-noise reliability provides information about the stability of reliability. The 95% CI is the mean signal-to-noise reliability ± (1.96*SE). The narrower the confidence interval, the less the mean signal-to-noise reliability estimate will change due to idiosyncratic features of specific plans. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the stability of reliability.
2. The distribution (minimum, 10th, 25th, 50th, 75th, 90th, maximum) of the plan-level signal-to-noise reliability estimates. Each plan's reliability estimate is a ratio of signal to noise, as described above [$\sigma2_{\text{plan-to-plan}} / (\sigma2_{\text{plan-to-plan}} + \sigma2_{\text{error}})$]. Variability between plans ($\sigma2_{\text{plan-to-plan}}$) is the same for each plan, while the specific plan error ($\sigma2_{\text{error}}$) varies. Reliability for each plan is an ordinal measure of how well one can determine where a plan lies in the distribution across plans, with higher estimates indicating better reliability. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the distribution of plan-level signal-to-noise reliability estimates. The number of plans in each stratum and the per-plan denominators of the performance rates are displayed in the summary tables.

This methodology allows us to estimate the reliability for each plan and summarize the distribution of these estimates.

2013 Submission

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009) in work produced for the National Committee for Quality Assurance (NCQA).

The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real

differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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

2020 Submission

Table 2 shows the point estimates of mean signal-to-noise reliability using the above methodology. The point estimate of mean signal-to-noise reliability ranges from 0.900 to 0.993 depending on the age group of patients captured in the performance rate.

Table 2. Point Estimates of Mean Signal-to-Noise Reliability by Product Type, Calendar Year 2019 Data

Appropriate Treatment for Upper Respiratory Infection	Point estimate: Mean Signal-To-Noise Reliability (Commercial)	Point estimate: Mean Signal-To-Noise Reliability (Medicaid)
Appropriate Treatment for Upper Respiratory Infection (Total)	0.983	0.992
Appropriate Treatment for Upper Respiratory Infection (age 3 Months-17 Years)	0.955	0.993
Appropriate Treatment for Upper Respiratory Infection (age 18-64)	0.980	0.979
Appropriate Treatment for Upper Respiratory Infection (age 65+)	0.900	0.945

Table 3 provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the URI (Total) measure for commercial and Medicaid plans overall and stratified by the denominator size (distribution of the number of eligible episodes of respiratory infection per plan). Over all commercial plans, the reliability estimate is 0.983, and the 95% CI is (0.978, 0.987), indicating very good reliability. Stratified analyses show that reliability increases as plan size gets larger and stays above 0.9. Over all Medicaid plans, the reliability estimate is 0.992 and the 95% CI is (0.989, 0.996), indicating very good reliability. Results from the stratified analyses show that reliability exceeds 0.9 for all terciles.

Table 3. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *URI (Total)* Measure by Terciles of the Denominator Size and for All Submissions Stratified by Plan Type, Calendar Year 2019 Data

Stratification	Number of Plans	Number of Eligible Members per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
All Commercial	**404**	**36 - 290,707**	**0.983**	**0.002**	**(0.978, 0.987)**
Tercile 1	133	36 – 2355	0.953	0.006	(0.941, 0.965)
Tercile 2	133	2364 – 9321	0.995	0.0002	(0.995, 0.996)
Tercile 3	138	9335 - 290707	0.999	0.0001	(0.998, 0.999)
All Medicaid	**222**	**99 – 231,499**	**0.992**	**0.002**	**(0.989, 0.996)**
Tercile 1	73	99 – 8760	0.985	0.004	(0.977, 0.993)
Tercile 2	73	8855 – 28798	0.998	0.0001	(0.998, 0.998)
Tercile 3	76	29,010 – 231,499	0.999	0.0001	(0.999, 0.999)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

Table 4 summarizes the distribution of plan-level signal-to-noise reliability estimates for the URI (Total) measure. Over all commercial plans, the estimates range from 0.616 to 1.0. The 50th percentile is 0.995, which exceeds the 0.70 threshold for reliability. For Medicaid plans, the estimates range from 0.698 to 1.0; the 10th percentile is 0.988, indicating very good reliability. This table also include the distribution of plan-level signal-to-noise reliability estimates stratified by denominator size. Reliability estimates are higher for plans with a larger denominator.

Table 4. Distribution of Plan-Level Signal-To-Noise Reliability for the URI (Total) Measure by Terciles of the Denominator Size and for All Submissions by Plan Type, Calendar Year 2019 Data

Distribution of Plan Estimates of Signal-to-Noise Reliability

Stratification	Number of Plans	Min	P10	P25	P50	P75	P90	Max
All Commercial	**404**	**0.616**	**0.966**	**0.988**	**0.995**	**0.999**	**0.999**	**1.0**
Tercile 1	133	0.613	0.867	0.962	0.980	0.987	0.990	0.995
Tercile 2	133	0.989	0.992	0.994	0.996	0.997	0.998	0.999
Tercile 3	138	0.996	0.997	0.998	0.999	0.999	1.0	1.0
All Medicaid	**222**	**0.698**	**0.988**	**0.996**	**0.999**	**0.999**	**1.0**	**1.0**
Tercile 1	73	0.770	0.976	0.990	0.996	0.997	0.998	1.0
Tercile 2	73	0.994	0.996	0.997	0.998	0.999	0.999	0.999
Tercile 3	76	0.997	0.999	0.999	0.999	1.0	1.0	1.0

2013 Submission

Reliability statistics for this measure were calculated using HEDIS health plan performance data for 2011. The results are as follows:

Commercial

URI - Reported rate 0.99

Medicaid

URI - Reported rate 1.00

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

2020 Submission

Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. Both plan types had an overall reliability greater than 0.70 indicating that the measure has very good reliability.

Overall commercial plans, the reliability estimate is 0.983, and the 95% CI is (0.978, 0.987), indicating very good reliability. Stratified analyses show that reliability increase as plan size gets larger and stay above 0.9. Overall Medicaid plans, the reliability estimate is 0.992 and the 95% CI is (0.989, 0.996), indicating very good reliability. Results from the stratified analyses show that reliability exceeds 0.9 for all terciles.

2013 Submission

NA

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

2020 Submission

Construct Validity Testing of Performance Measure Score

NCQA performs Pearson correlation for construct validity using HEDIS health plan data. The test estimates the strength of linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a strong positive linear association: an increase in values of one variable is associated with increase in value of another variable. A value of 0 indicates no linear association. A value of -1 indicates a strong negative relationship in which an increase in values of the first variable is associated with a decrease in values of the second variable. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We adjusted our p-values to account for testing multiple correlations and used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

We tested for construct validity by exploring the following:

- Is *Appropriate Treatment for Upper Respiratory Infection* positively correlated with the HEDIS *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* measure which assesses the percentage of episodes for members ages 3 months and older with a diagnosis of acute bronchitis/bronchiolitis that did not result in an antibiotic dispensing event?

- Is *Appropriate Treatment for Upper Respiratory Infection* positively correlated with the HEDIS *Use of Imaging Studies for Low Back Pain* measure which assesses members with a primary diagnosis of low back pain who did not have an imaging study within 28 days of the diagnosis?

- Is *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* negatively correlated with the HEDIS *Antibiotic Utilization* measure which assesses the average number of outpatient antibiotic prescriptions per member per year?

We hypothesized that health plans with a high rate for the *Appropriate Treatment for Upper Respiratory Infection* measure would have a high rate for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* measure. Acute bronchitis/bronchiolitis and upper respiratory infections are largely driven by viruses and antibiotic stewardship efforts such as patient education and prior authorization target these common outpatient diagnoses to avoid inappropriate antibiotic prescribing.

We hypothesized that health plans with a high rate for the *Appropriate Treatment for Upper Respiratory Infection* measure would have a high rate for the *Use of Imaging Studies for Low Back Pain* measure. Both measures assess the rate of avoiding potentially inappropriate care (e.g. providing antibiotics for viral infections or performing unnecessary imaging). Health plans that are committed to protecting patients from undue harm and reducing costs by avoiding unnecessary medications and tests should perform well on both measures.

We hypothesized that health plans with a high rate for the *Appropriate Treatment for Upper Respiratory Infection* measure would have a low rate for the *Antibiotic Utilization* measure. At least 30% of outpatient antibiotic prescriptions are inappropriate and treatment for acute bronchitis/bronchiolitis is a large contributor to broad outpatient antibiotic overuse (CDC, 2019). Health plans with higher rates for appropriate antibiotic prescribing for upper respiratory infection should have lower overall antibiotic utilization.

Centers for Disease Control and Prevention. (2019). Antibiotic Use in the United States, 2018 Update: Progress and Opportunities. Atlanta, GA: US Department of Health and Human Services, CDC; 2019.

Systematic Assessment of Face Validity of Performance Measure Score

NCQA develops measures using a standardized process described below.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical measurement advisory panels (MAPs), whose members are authorities on clinical priorities for measurement, participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness, and feasibility. This information is gathered into a work-up format, which is vetted by the MAPs, the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. At this step, face validity is systematically determined by the CPM, which uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment. For the most recent updates to this measure in January 2019, all members of the CPM voted to approve moving forward with the proposed changes.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA about proposed new measures. Public comment offers an opportunity to assess the validity, feasibility, importance and other attributes of a measure from a wider audience. For this measure, a majority of public comment respondents supported the measure. NCQA MAPs and the technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. Face validity is then again systematically assessed by the CPM. The CPM reviews all comments before making a final decision and votes to recommend approval of new measures for HEDIS. NCQA's Board of Directors then approves new measures.

2013 Submission

NCQA identified and refined measure management into a standardized process called the HEDIS measure life cycle. The following steps outline the components of the life cycle that are used to ensure that measure testing adheres to the highest standard possible.

***Step 1:** Topic selection is the process of identifying measures that meet criteria consistent with the overall model for performance measurement. There is a huge universe of potential performance measures for future versions of HEDIS. The first step is identifying
measures that meet formal criteria for further development.

NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. Refer to What Makes a Measure "Desirable?" The work-up is vetted by NCQA's MAPs, the TMAP, and various other panels.

***Step 2:** Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic
selection phase.

Development includes the following tasks:

1. Ensure funding throughout measure testing
2. Prepare a detailed conceptual and operational work-up that includes a testing proposal
3. Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures

The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

***Step 3:** Public Comment is a 30-day period of review that allows interested parties to offer feedback to the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM will be included in the next HEDIS year and reported as first-year measures.

***Step 4:** First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's Quality Compass or in accreditation scoring.

The first-year distinction guarantees that a measure can be efficiently collected, reported and audited before it is used for public accountability or accreditation. The purpose of this first-year distinction is to ensure that there are no unforeseen problems when the measure is implemented in the real world.

NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

***Step 5:** Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be reported in Quality Compass and may be used for scoring in accreditation.

***Step 6:** Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reevaluated at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments contribute to measure evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, a third of the measurement set is researched for changes in clinical guidelines or health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be

recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.
What makes a measure "Desirable"? Whether considering the value of a new measure or the continuing worth of an existing one, we must define what makes a measure useful. HEDIS measures encourage improvement. The defining question for all performance measurement— "Where can measurement make a difference?"— can be answered only after considering many factors. NCQA has established three areas of desirable characteristics for HEDIS measures, discussed below.

1. Relevance: Measures should address features that apply to purchasers or consumers, or which will stimulate internal efforts toward quality improvement. More specifically, relevance includes the following attributes.

Meaningful: What is the significance of the measure to the different groups concerned with health care? Is the measure easily interpreted? Are the results meaningful to target audiences? Measures should be meaningful to at least one HEDIS audience (e.g., individual consumers, purchasers or health care systems). Decision makers should be able to understand a measure's clinical and economic significance. Important to health: What is the prevalence and overall impact of the condition in the U.S. population? What significant health care aspects will the measure address? We should consider the type of measure (e.g., outcome or process), the prevalence of medical condition addressed by the measure and the seriousness of affected health outcomes. Financially important: What financial implications result from actions evaluated by the measure? Does the measure relate to activities with high financial impact? Measures should relate to activities that have high financial impact. Cost effective: What is the cost benefit of implementing the change in the health care system? Does the measure encourage the use of cost-effective activities or discourage the use of activities that have low cost-effectiveness? Measures should encourage the use of cost-effective activities or discourage the use of activities that have low cost-effectiveness.
Strategically important: What are the policy implications? Does the measure encourage activities that use resources efficiently? Measures should encourage activities that use resources most efficiently to maximize member health.
Controllable: What impact can the organization have on the condition or disease? What impact can the organization have on the measure? Health care systems should be able to improve their performance. For outcome measures, at least one process should be controlled and have an important effect on outcome. For process measures, there should be a strong link between the process and desired outcome.
Variation across systems: Will there be variation across systems? There should be the potential for wide variation across systems.
Potential for improvement: Will organizations be able to improve performance? There should be substantial room for performance improvement.

2. Scientific soundness: Perhaps in no other industry is scientific soundness as important as in health care. Scientific soundness must be a core value of our health care system—a system that has extended and improved the lives of countless individuals.

Clinical evidence: Is there strong evidence to support the measure? Are there published guidelines for the condition? Do the guidelines discuss aspects of the measure? Does evidence document a link between clinical processes and outcomes addressed by the measure? There should be evidence documenting a link between clinical processes and outcomes.
Reproducible: Are results consistent? Measures should produce the same results when repeated in the same population and setting. Valid: Does the measure make sense? Measures should make sense logically and clinically, and should correlate well with other measures of the same aspects of care.
Accurate: How well does the measure evaluate what is happening? Measures should precisely evaluate what is actually happening.
Risk adjustment: Is it appropriate to stratify the measure by age or another variable? Measure variables should not differ appreciably beyond the health care system's control, or variables should be known and measurable. Risk stratification or a validated model for calculating an adjusted result can be used for measures with confounding variables.

Comparability of data sources: How do different systems affect accuracy, reproducibility and validity? Accuracy, reproducibility and validity should not be affected if different systems use different data sources for a measure.

3. Feasibility: The goal is not only to include feasible measures, but also to catalyze a process whereby relevant measures can be made feasible.

Precise specifications: Are there clear specifications for data sources and methods for data collection and reporting? Measures should have clear specifications for data sources and methods for data collection and reporting.

Reasonable cost: Does the measure impose a burden on health care systems? Measures should not impose an inappropriate burden on health care systems.

Confidentiality: Does data collection meet accepted standards of member confidentiality? Data collection should not violate accepted standards of member confidentiality. Logistical feasibility: Are the required data available? Auditability: Is the measure susceptible to exploitation or "gaming" that would be undetectable in an audit? Measures should not be susceptible to manipulation that would be undetectable in an audit.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Statistical results of construct validity testing

Table 5. Results of Pearson Correlation Coefficient for Commercial and Medicaid health plans for the Appropriate Treatment for Upper Respiratory Infection (Total) Measure, Calendar Year 2019 Data

Measure	Avoidance of Antibiotic Treatment for Acute Bronchitis/ Bronchiolitis (Total)	Antibiotic Utilization - Average Scrips for Antibiotics PMPY (M/F)	Use of Imaging Studies for Low Back Pain (Total)
Commercial	0.68	-0.74	0.622
(N=, p value =)	(396, p < 0.001)	(392, p < 0.001)	(392, p < 0.001)
Medicaid	0.68	-0.73	0.41
(N=, p value =)	(212, p < 0.001)	(194, p < 0.001)	(208, p < 0.001)

Results of face validity assessment

Input from our multi-stakeholder measurement advisory panels and those submitting to public comment indicate the measure has face validity. The CPM unanimously supported the measure.

2013 Submission

Step 1: The Appropriate treatment for children with upper respiratory infection (URI) measure was developed to address a gap in care concerning the need to decrease excess antibiotic use in ambulatory practice, fueled by the epidemic increase in antibiotic resistant Streptococcus pneumonia. NCQA's Performance Measurement Department and the URI/Bronchitis Technical Subgroup worked together to determine the most appropriate way to meet this objective.

Step 2: The measure was written, field-tested, and presented to the CPM in 2001. The CPM recommended sending the measure to public comment.

Step 3: The measure was released for Public Comment in spring 2002. We received and responded to comments on this measure. The CPM recommended moving this measure to first year data collection. The voting process involved a simple majority vote with a quorum of CPM members.

Step 4: The Appropriate treatment for children with upper respiratory infection (URI) measure was introduced in HEDIS 2003. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting. The voting process involved a simple majority vote with a quorum of CPM members.

Step 5: The Appropriate treatment for children with upper respiratory infection (URI) measure will be reevaluated in 2013.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2020 Submission

Interpretation of construct validity testing

Commercial:

Appropriate Treatment for Upper Respiratory Infection is positively correlated with *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* (correlation coefficient = 0.68, p < 0.001) and *Use of Imaging Studies for Low Back Pain* (correlation coefficient = 0.62, p < 0.001). It is negatively correlated with *Antibiotic Utilization* (correlation = -0.74, p < 0.001).

Medicaid:

Appropriate Treatment for Upper Respiratory Infection is positively correlated with *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* (correlation coefficient = 0.68, p < 0.001) and *Use of Imaging Studies for Low Back Pain* (correlation coefficient = 0.41, p < 0.001). It is negatively correlated with *Antibiotic Utilization* (correlation = -0.73, p < 0.001).

2013 Submission

NA

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2020 Submission

Exclusion data is not collected during annual HEDIS reporting. Exclusions were assessed during development of the measure. Initial measure field testing was conducted across three national health plan organizations reflecting all lines of business (Medicare, Medicaid and commercial) using their 2016 enrollment and claims data. The raw frequency of exclusion events and proportion of denominator events excluded were evaluated. The steps to conduct this testing are described below:

1. NCQA recruited three health plan organizations with Medicare, Medicaid and commercial product lines to participate in field testing. These sites provided relevant data on their member population as well as qualitative information on their experience of collecting and reporting antibiotic use information.
2. The NCQA team developed a standardized data collection protocol based on a uniform data model developed for the specific purpose of collecting standardized, electronic clinical data. Each plan was asked to submit an aggregate table of overall plan descriptive information as well as a member-level comma-separated value (csv) file containing all the requested elements of the data model.
3. Using the csv file submitted by each plan, NCQA identified the eligible member population.
4. Among the eligible population, denominator events for each plan were identified following the logic:

   Step 1

   Identify all members who had an outpatient visit (Outpatient Value Set), a telephone visit (Telephone Visits Value Set), an online assessment e-visit or virtual check-in (Online Assessments Value Set) an observation visit (Observation Value Set) or an ED visit (ED Value Set) during the Intake Period, with a diagnosis of URI (URI Value Set).

   Step 2

   Determine all URI Episode Dates. For each member identified in step 1, determine all outpatient, telephone, observation or ED visits, e-visits and virtual

check-ins with a URI diagnosis.

Exclude visits that result in an inpatient stay (Inpatient Stay Value Set).

5. Among the denominator events, NCQA evaluated the frequency of each exclusion and the proportion of total denominator events removed when implementing the exclusion independently.

The hospice exclusion is not tested individually, but rather implemented in all HEDIS measures based on expert panel feedback on clinical appropriateness.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2020 Submission

Table 6. Exclusion Analysis: Appropriate Treatment for Upper Respiratory Infection

Exclusion	Line of Business	Plan A: Total No. Visits	Plan A: No. Excluded	Plan A: Prop. Excluded	Plan B: Total No. Visits	Plan B: No. Excluded	Plan B: Prop. Excluded	Plan C: Total No. Visits	Plan C: No. Excluded	Plan C: Prop. Excluded
Competing Diagnosis (1)	**Commercial**	--	--	--	--	--	--	38,875	9,989	25.70%
Competing Diagnosis (2)	**Medicaid**	42922	7,720	18.00%	68,269	13,145	19.30%	46,559	11,921	25.60%
Competing Diagnosis (3)	**Medicare**	2071	300	14.50%	--	--	--	5,678	728	12.80%
Comorbid Condition (1)	**Commercial**	--	--	--	--	--	--	38,875	413	1.10%
Comorbid Condition (2)	**Medicaid**	42,922	1,155	2.70%	68,269	719	1.10%	46,559	385	0.80%
Comorbid Condition (3)	**Medicare**	2,071	326	15.70%	--	--	--	5,678	318	5.60%
Negative Medication History (1)	**Commercial**	--	--	--	--	--	--	38,875	3,248	8.40%
Negative Medication History (2)	**Medicaid**	42,922	3,545	8.30%	68,269	6,978	10.20%	46,559	5,175	11.10%
Negative Medication History (3)	**Medicare**	2,071	254	12.30%	--	--	--	5,678	890	15.70%

-- cell intentionally left blank

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* **Note***: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2020 Submission

In order to capture unique outpatient, telephone, online, observation or ED visits for upper respiratory infection that did not result in an antibiotic prescribing event, it is necessary to exclude denominator events that are confounded by proximal outpatient events for competing diagnoses where an antibiotic is warranted (i.e. competing diagnosis such as pharyngitis). Additionally it is necessary to exclude denominator events that are confounded by previously ongoing antibiotic treatment within the prior 30 days as an antibiotic prescription during the window of evaluation may not be attributed to the denominator event for upper respiratory infection, but instead a refill event for chronic antibiotic use or continued treatment for a previous infection (i.e. negative medication history). The competing diagnosis exclusion accounted for 25.7% of commercial events, ranged from 18% to 25.6% for Medicaid and ranged from 12.8% to 14.5% for Medicare. The negative antibiotic medication exclusion accounted for 8.4% of commercial events, ranged from 8.3% to 11.1% for Medicaid and ranged from 12.3% to 15.7% for Medicare. The Centers for Disease Control and Prevention recommends against antibiotic prescribing for upper respiratory infections and note different treatment considerations should be given to individuals with select immunocompromising comorbidities (i.e. comorbidity such as cystic fibrosis). Individuals with other comorbidities potentially requiring antibiotics should be excluded from the measure and testing indicated that these accounts for 1.1% of commercial events, ranged from 0.8% from 2.7% for Medicaid and ranged from 5.6% to 15.7% for Medicare.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2020 Submission

N/A. Not an intermediate or health outcome, PRO-PM, or resource use measure.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2020 Submission

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standard error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than 0.05, then the two plans' performance is significantly different from each other.

2013 Submission

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.

The analysis included a geographically diverse universe of commercial and Medicaid plans between 2009 and 2011.

Number of commercial health plans, 2009= 424

Number of commercial health plans, 2010= 421

Number of commercial health plans, 2011= 405

Number of Medicaid health plans, 2009= 132

Number of Medicaid health plans, 2010= 142

Number of Medicaid health plans, 2011= 154

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2020 Submission

Table 7 summarizes the distribution of plan-level performance for the URI (Total) measure, for Commercial and Medicaid Plans. There is an 0.11 (i.e. 11 percentage points) gap in performance between Commercial plans at the 25th and 75th percentiles, and a 0.06 percentage point gap in performance among Medicaid plans. The difference in performance between plans in the 25th percentile and 75th percentile is statistically significant.

Table 7. Variation in Performance for Commercial and Medicaid health plans for the Appropriate Treatment for Upper Respiratory Infection (Total) Measure, Calendar Year 2019 Data

Measures	N	Mean	Min	P10	P25	Median	P75	P90	Max	IQR	P value
Commercial	404	0.79	0.49	0.68	0.74	0.80	0.85	0.89	0.96	0.11	< 0.001
Medicaid	222	0.87	0.50	0.79	0.85	0.88	0.91	0.94	1.0	0.06	< 0.001

N = Number of plans reporting

IQR = Interquartile range

p-value = p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2013 Submission

Commercial

2011 RATE

N 405

Mean 84.49

StdDev 8.02

Min 44.48

P10 74.85

P25 80.25

P50 86.07

P75 89.74

Max 98.48

2010 RATE

N 421

Mean 83.40

StdDev 8.65

Min 31.06

P10 73.23

P25 79.02

P50 85.01

P75 88.98

Max 100.00

2009 RATE

N 424
Mean 83.61
StdDev 8.37
Min 42.00
P10 72.86
P25 78.93
P50 84.96
P75 89.20
Max 100.00
Medicaid
2011 RATE
N 154
Mean 87.18
StdDev 6.07
Min 72.24
P10 79.24
P25 83.39
P50 87.49
P75 91.86
Max 98.87
2010 RATE
N 142
Mean 86.04
StdDev 6.52
Min 70.02
P10 77.68
P25 82.12
P50 85.78
P75 90.65
Max 98.30
2009 RATE
N 132
Mean 85.49
StdDev 6.84
Min 59.21
P10 78.09
P25 81.12
P50 85.61
P75 91.11
Max 98.46

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities?
(i.*e., what do the results mean in terms of statistical and meaningful differences?*)

2020 Submission

There is an 0.11 gap in performance for commercial plans between the 25th and 75th percentile for commercial plans, representing about 494 lives for the median plan (median plan denominator = 4,493). There is a 0.06 gap in performance for Medicaid plans, representing about 1,140 lives for the median plan (median plan denominator = 19,004). In each instance, the gap is statistically significant, and represents a meaningful difference between lower and higher performing plans, demonstrating opportunity for improvement.

2013 Submission

The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. It includes number of health plans, percentiles, mean, min, max and standard deviations.

Data is summarized at the health plan level (i.e. "N" represents the number of health plans). The rate is reported as an inverted rate (i.e. 1- numerator/denominator) to reflect the number of people in the health plans that were not dispensed an antibiotic.

Data is stratified by year and product line (i.e. commercial, Medicare, Medicaid).

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

This measure has only one set of specifications.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

2020 Submission

HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:

- Information practices and control procedures

- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

2013 Submission

NA

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse;* **if no empirical sensitivity analysis***, identify the approaches for handling missing data that were considered and pros and cons of each*)

2020 Submission

HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the feasibility of the measure when widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

2013 Submission

NA

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted;* **if no empirical analysis***, provide rationale for the selected approach for missing data*)

2020 Submission

All of the commercial and Medicaid health plans reporting for the HEDIS 2020 (CY2019) measurement year were audited as described above. This means that the auditors did not find missing data sources for any of the health plan data submissions and determined that none of the rates were materially biased.

2013 Submission

NA

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the managed care organization´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

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

In addition to the HEDIS audit, NCQA provides a system to allow “real-time” feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system informs both annual updates to the measures as well as routine

re-evaluation of measures. These processes include updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

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

Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Quality Improvement (Internal to the specific organization)	Public Reporting NCQA Health Plan Ratings https://www.ncqa.org/hedis/reports-and-research/ratings-2020/ NCQA Annual State of Health Care Quality https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report CMS QRS program https://www.cms.gov/files/document/2021-qrs-measure-technical-specifications.pdf Public Health/Disease Surveillance Centers for Disease Control and Prevention (CDC) Measuring Outpatient Antibiotic Prescribing https://www.cdc.gov/antibiotic-use/community/programs-measurement/measuring-antibiotic-prescribing.html Payment Program Quality Payment Program https://qpp.cms.gov/mips/explore-measures Regulatory and Accreditation Programs NCQA Health Plan Accreditation https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/ Quality Improvement (external benchmarking to organizations) NCQA Quality Compass

Specific Plan for Use	Current Use (for current use provide URL)
	https://www.ncqa.org/programs/data-and-information-technology/data-purchase-and-licensing/quality-compass/

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

NCQA HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan ratings which are reported on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. Due to COVID-19, NCQA will not release 2010-2021 Health plan ratings for any product line. However, in 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
NCQA STATE OF HEALTH CARE QUALITY: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care.
QUALIFIED HEALTH PLAN (QHP) QUALITY RATING SYSTEM (QRS): This measure is used in the Qualified Health Plan (QHP) Quality Rating System (QRS) which provides comparable information to consumers about the quality of health care services and QHP enrollee experience offered in the Marketplaces.
CDC MEASURING OUTPATIENT ANTIBIOTIC PRESCRIBING: Monitoring of outpatient antibiotic prescribing data is regularly conducted to analyze national and state antibiotic prescribing data in order to better understand trends in outpatient antibiotic prescribing, to identify where interventions to improve prescribing are most needed, and to measure progress. The CDC website lists average national performance on the URI HEDIS measure. The CDC website are publicly available to all audiences.
QUALITY PAYMENT PROGRAM:
The Quality Payment Program (QPP) is a quality and cost incentive program that uses payment adjustments to promote high quality and high value care delivery by eligible clinicians (EC). QPP provides performance-based payment adjustments to ECs, both negative and positive, for services furnished to Medicare Part B beneficiaries. EC performance is graded on quality measure performance, cost of care, engagement in clinical practice improvement activities, and use of Certified EHR Technology (CEHRT). Performance can be reported at the individual (clinician) or group (practice) level. In 2017, 1,006,319 ECs participated in MIPS, representing 95% of all eligible clinicians across the 50 states. 54% participated as a part of a group, 12% as individual clinicians, and 34% as a part of an Advanced Payment Model.
NCQA HEALTH PLAN ACCREDITATION: This program is a widely recognized, evidence-based program dedicated to quality improvement and measurement. It provides a comprehensive framework for organizations to align and improve operations in areas that are most important to states, employers and consumers. It's the only evaluation program that bases results on actual measurement of clinical performance (HEDIS®?measures) and consumer experience (CAHPS®?measures). As of October 2020, there are 507 commercial, 228 Medicare and 178 Medicaid health plans with accreditation, representing entities from all states and geographic regions.
NCQA QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting health plans, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference (now the Quality Innovation Series), NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly and insight into new measure development projects. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

NCQA measures are evaluated regularly. During this "reevaluation" process, we seek broad input on the measure, including input on performance and implementation experience. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, 30 day public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

NCQA released proposed measurement changes in our annual HEDIS Public Comment period in 2019, which was available to all audiences to provide feedback on proposed measure updates and changes. Advisory panels of experts in antibiotic overuse and infectious diseases were also consulted.

Proposed changes included the addition of the Medicare product line, the expansion of the age group to include members 3 months of age and older, and the transition to an episode-based denominator.

4a2.2.2. Summarize the feedback obtained from those being measured.

In general, health plans have not reported significant barriers to implementing this measure, as it uses the administrative data collection method. A review of questions submitted by customers through NCQA's Policy Clarification System over the past 3 years shows that the majority of inquiries centered around minor clarification of the specification, such as whether specific scenarios/patients met qualifications for inclusion in the measure's eligible population. NCQA responded to all questions to ensure consistent implementation of the specifications.

4a2.2.3. Summarize the feedback obtained from other users

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs, such as the Quality Rating System and Merit-based Incentive Payment System.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

NCQA considers feedback from the public, experts and other stakeholders when making decisions about updating measure specifications. As a result of the feedback we received, NCQA expanded the age range of the measure and reporting to all three product lines.

Improvement

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

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Given significant changes to the measure denominator and age ranges covered between MY 2018 and MY 2019, trends in performance cannot be assessed. For MY 2019, the average percentage of episodes that were not associated with a dispensed antibiotic was 78.6% for commercial plans and 86.9% for Medicaid plans (full performance distribution details in section 1b). These proportions indicate high health plan performance on antibiotic prescribing for upper respiratory infection and substantiates continued use of the measure, particularly for commercial plans. With a national focus on antibiotic stewardship, the goal is for health plans to continue driving progress in appropriate and conservative antibiotic use.

4b2. Unintended Consequences

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

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

There were no identified unintended consequences for this measure during testing or since implementation.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

There were no identified unintended benefits for this measure during testing or since implementation.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0058 : Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (AAB)

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

N/A

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

Both measure specifications focus on inappropriate antibiotic prescribing. The current measures considers antibiotic prescribing in the case of upper respiratory infections, while NQF #0058 considers prescribing in the case of acute bronchitis or bronchiolitis. The diagnosis may impact clinician decision for antibiotic prescribing.

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

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

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Brittany, Wade, wade@ncqa.org, 202-530-0463-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

The Antibiotic Overuse Measurement Advisory Panel advised NCQA during the measure's recent reevaluation. They evaluated the measure specifications, assessed the face validity of measures, and reviewed field test results. The Advisory Panel consisted of a balanced group of experts, including representatives from health plans and specialty societies. In addition, NCQA also vetted these measures with a host of other stakeholders, including the Committee on Performance Measurement (CPM), a voting body that reviews evidence and input from stakeholders in order to assess a measure's appropriateness for NCQA programs.

Antibiotic Overuse Measurement Advisory Panel (AOMAP)

Diana Buist, MPH, PhD, Kaiser Permanente Washington Health Research Institute

Jonathan Finkelstein, MD, MPH, Boston Children's Hospital

Jeffrey Gerber, PhD, The Children's Hospital of Philadelphia

Catherine Gillespie, PhD, MPH, AARP Public Policy Institute

Jeffrey Linder, MD, MPH, Northwestern University

Karl Madaras-Kelly, PharmD, PMH, Idaho State University

Rita Mangione-Smith, MD, MPH, University of Washington

Dat Tran, MD, Oregon Public Health Division

Committee on Performance Measurement (CPM)

Andy Baskin, MD, CVS Health/Aetna

Elizabeth Drye, MD, SM, Yale School of Medicine

Mark Friedberg, MD,MPP, Blue Cross Blue Shield of Massachusetts

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine S. Hunter, MD, RADM, MC, USN, Self-employed, Independent Board Director

David Kelley, MD, MPA, Chief Medical Officer, Pennsylvania Department of Human Services

Jeff Kelman, MD, MMSc., Chief Medical Officer, Center for Medicare Department of Health and Human Services (DHHS)

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Self Employed

Amanda Parsons, MD, MBA, MetroPlus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

Rudy Saenz, MD, MMM, GACOG, Riverside Medical Clinic

Marcus Thygeson, MD, MPH, Bind Benefits

JoAnn Volk, MA, Georgetown University, Center on Health Insurance Reforms

Rose Baez, RN, MSN, MBA, CPHQ, Blue Cross Blue Shield

Jeff Brady, MD, MPH, AHRQ

Ron Kline, MD, Office of Personnel Management

Danielle Lloyd, MPH, America's Health Insurance Plan (AHIP)

Chelsey Richards, MD, MPH, FACP, Centers for Disease Control and Prevention

Anecia Suneja, CNS-BC, Veterans Health Administration (VHA)

Sheri Winsper, RN, MSN, MSHA, National Quality Forum (NQF)

Measure Developer/Steward Updates and Ongoing Maintenance

The URI Technical Subgroup advised NCQA during measure development. They evaluated the way staff specified the measure, reviewed field test results, and assessed NCQA's overall desirable attributes of Relevance, Scientific Soundness, and Feasibility. The advisory panel consisted of a balanced group of experts. In addition to this advisory panel, we vetted the measure with a host of other stakeholders, as is our process. Thus, our measures are the result of consensus from a broad and diverse group of stakeholders.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2004

Ad.3 Month and Year of most recent revision: 05, 2019

Ad.4 What is your frequency for review/update of this measure? Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

Ad.5 When is the next scheduled review/update for this measure? 12, 2021

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.