NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0577

Measure Title: Use of Spirometry Testing in the Assessment and Diagnosis of COPD

Measure Steward: National Committee for Quality Assurance

Brief Description of Measure: This measure assesses the percentage of patients 40 years of age and older with a new diagnosis of COPD or newly active COPD, who received appropriate spirometry testing to confirm the diagnosis.

Developer Rationale: This measure assesses whether patients considered to have a diagnosis of COPD through the presence of symptoms and risk factors (e.g., dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease) had a spirometry assessment to confirm the diagnosis. The improvement in quality envisioned by the use of this measure is to ensure that patients receive spirometry testing to confirm a COPD diagnosis and determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to guide therapy.

Numerator Statement: The number of patients with at least one claim/encounter for spirometry during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date.

Denominator Statement: All patients age 42 years or older as of December 31 of the measurement year, who had a new diagnosis of COPD or newly active COPD during the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year.

Denominator Exclusions: This measure excludes patients who use hospice services, and those with nonacute inpatient stays.

Measure Type: Process

Data Source: Claims

Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Dec 04, 2009 **Most Recent Endorsement Date:** Aug 03, 2016

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 U

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**

Evidence Summary or **Summary of prior review in 2016**

- During prior years, the Committee agreed the evidence was appropriate and consistent for the use of spirometry to confirm the diagnosis of Chronic Obstructive Pulmonary Disease (COPD).
- During the 2016 review, the developer provided the 2015 GOLD guidelines and the 2011 American College of Physicians COPD clinical guidelines.
  - Dated 2015, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines provide ungraded recommendations for COPD assessment. The GOLD guidelines referenced 613 studies to update the previous set of guidelines from 2013. The recommendation for spirometry to confirm a COPD diagnosis was based on a systematic review and three observational studies.
  - The American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society 2011 guidelines graded the evidence as a strong recommendation with moderate quality evidence (the second highest ranking in this grading system). The recommendation is: "spirometry should be obtained to diagnose airflow obstruction in patients with respiratory symptoms." The recommendation for spirometry to confirm a COPD diagnosis cited 17 randomized control trials, meta-analyses, systematic reviews, and observational studies.
  - The Institute for Clinical Systems Improvement (ICSI) Guidelines, dated 2013, referenced a systematic review related to spirometry testing to confirm a COPD diagnosis. No grading was provided

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:

- This process measure assesses the percentage of patients 40 years of age and older with a new diagnosis of COPD or newly active COPD, who received appropriate spirometry testing to confirm the diagnosis.
- The developer provided a logic model demonstrating that appropriate diagnosis of COPD would guide care resulting in decreased frequency and severity of exacerbations, urgent care and emergency department visits and inpatient hospital stays.
- The developer provided the updated 2020 GOLD guidelines.
    - The 2020 GOLD guidelines for COPD recommends spirometry for confirmation of COPD diagnosis. Spirometry is the most reliable measurement of airflow obstruction. No grading is assigned to statements or associated evidence referenced in the GOLD guidelines.
- Per the developer, 2020 GOLD guidelines are supported by two observational studies, as well as a summary article from an American Thoracic Society/European Respiratory Society (ATS/ERS) series on interpretative strategies for lung function tests. The developer noted ICSI retired its guidelines and endorsed the Veteran's Affairs and Department of Defense (VA/DOD) Clinical Practice Guideline for the Management of COPD.
    - The developer also referenced the Veteran's Affairs and Department of Defense Clinical Practice Guideline for the Management of COPD. The developer noted the Va/DOD rated this recommendation as Strong. The VaDOD recommendation states "We recommend that spirometry, demonstrating airflow obstruction (post-bronchodilator forced expiratory volume in one second/forced vital capacity [FEV1/FVC] <70%, with age adjustment for more elderly individuals), be used to confirm all initial diagnoses of chronic obstructive pulmonary disease (COPD) (Strong For)." Strong recommendation indicates the Va/DOD workgroup is confident that the desirable consequences outweigh the undesirable consequences. The developer provided a Quantity, Quality and Consistency of guideline. The developer noted that evidence is based on strong recommendation for spirometry in 2 systematic reviews, 1 RCT, 3 case series and 1 cohort trial.

Exception to evidence
N/A

Questions for the Committee:

- The evidence provided by the developer is updated, and directionally the same compared to that for the previous NQF review.
- For structure, process, and intermediate outcome measures:
    - 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

Per page 15 of the 2019 NQF Measure Evaluation Criteria – Clinical Evidence Algorithm: 1. Measure assesses a process → 3. It is based on a systematic review and grading → 4. Summary of QQC included → 5b. Quantity: Low/High; Quality: High; Consistency: High >>> Rate as moderate

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.

- The developer provided performance data from HEDIS database on patients who were considered to have a new diagnosis of COPD and had spirometry testing to confirm the diagnosis. The data was stratified by year and commercial, Medicaid or Medicare insurance.
    - Commercial (2016, 2017, 2018): mean (40.5-41.7%); SD (7.1-7.3%); Interquartile Range (7.2-8.4%) Performance at 90th percentile (49.7-49.8%)
    - Medicare (2016, 2017, 2018): mean (34.0-34.9%); SD (9.8-10.7%); Interquartile Range (11.2-11.8%) Performance at 90th percentile (45.2-48.1%)
    - Medicaid (2016, 2017, 2018): mean (31.1-31.7%); SD (8.7-9.3%); Interquartile Range (9.6-11.4%) Performance at 90th percentile (41.1-44.6%)
- The variations and low performance indicate gap in care and an opportunity for improvement.

Disparities

- Developer does not currently collect performance data stratified by race, ethnicity, or language.
- The developer noted literature that women and African Americans are more likely to not have a COPD diagnosis at all stages of airflow obstruction.
- The developer also reviewed literature that indicates the rate of spirometry testing was higher in Asian/Pacific Islander patients and Black patients with a new diagnosis of COPD or newly active COPD compared to White patients and Hispanic patients were found to be less likely than White patients to receive spirometry.

Questions for the Committee:

- Specific questions on information provided for gap in care.
- Is there a gap in care that warrants a national performance measure?
- If no 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

Comments:

** This is a process measure for which the underlying rationale and high quality evidence has not changed substantially since last NQF endorsement

** This continuing guideline looks at the frequency of spirometry being performed with or soon after the diagnosis of COPD is made. Spirometry is a "stable" process in that there is accepted standardization of technique and the means of measurement is relatively simple and relatively inexpensive. The evidence cited of new and ongoing systematic clinical reviews continue to support that the process of spirometry is necessary for the adequate diagnosis of COPD (both in rule in and rule out situations) and that spirometry provides useful prognostic information for COPD and its treatments.

** Evidence is updates and in same direction as previous submission. There was the new GOLD guidelines and sprometry as the most valuable confirmation for COPD.

** This is a maintenance measure in which additional evidence is provided that relates directly to the outcome.

**I'm not aware of any unmentioned studies that aren't cited in the submission. The most robust updated evidence provided is the 2020 GOLD guidelines

1b. Performance Gap

Comments:

** YES. There is opportunity for improvement

** The ideal performance measure might track the recording of a spirometry for those with a history of the symptoms of COPD (dyspnea, chronic cough or sputum, and history of risk factor exposure) but symptoms are not coded. For the years of HEDIS results provided, the gap persists in the process of patients given a diagnosis of COPD having spirometry recorded (in a reasonably liberal period of time).

** With the Mean percentage for all payer groups for the 90th percentile between 41.1-49.8%, it appears the is a performance gap.

** Current performance data is provided that shows a significant gap in care.

**Yes, I think it does

Disparities:

Comments:

** Disparity data not collected by developer; however, there is evidence in published literature supporting the contention that there are ongoing disparities.

** The discussion provides a succinct explanation of the limitations of current data collections to determine race, ethnicity, and socioeconomic status and of trying to extrapolate from data that is collected to assume someone's race, ethnicity, or socioeconomic status.

** It appears there are disparities, but this measure does not consider these data.

**There were disparities evident which the developer addressed adequately.

**Data isn't provided, although the developer mentions disparities are likely

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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.

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

Evaluators: Primary Care and Chronic Illness project team staff
Evaluation of Reliability and Validity: Link A (Project Team staff)

Reliability:

- During the 2012 review NCQA provided reliability statistics for this measure using HEDIS health plan performance data for 2010. The beta-binomial method was used to determine the ratio of signal to noise with results of 0.94 for Commercial, 0.92 for Medicaid and 0.97 for Medicare.
- In the 2016 review NCQA provided measure score reliability and construct validity using data from all health plans that submitted HEDIS data for this measure in 2012 and 2015. These analyses included all of the health plans (365 Commercial, 355 Medicare, and 124 Medicaid).
- NCQA provided updated reliability scores calculated using HEDIS health plan performance data for 2018.
- Beta-Binomial Statistics:

Commercial			Medicare			Medicaid		
Avg.	Overall	10th-90th	Avg.	Overall	10th-90th	Avg.	Overall	10th-90th
0.76	0.91	0.48-0.96	0.90	0.98	0.72-0.99	0.88	0.95	0.66-0.98

- For signal to noise, 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:

- During the previous NQF review in 2014, developer provided the below Construct Validity results:
  - The developer hypothesized that organizations that perform well on the measure should perform well on other similar HEDIS measures.
  - Pearson correlation coefficients were calculated to estimate the strength of the association of this spirometry measure to two other COPD measures:
    - Pharmacotherapy Management of COPD Exacerbation: Bronchodilator Indicator and
    - Pharmacotherapy Management of COPD Exacerbation: Systemic Corticosteroid Indicator.
  - Results are reported within a ranges from -1 and +1 indicated that the COPD measures were significantly ($p<.05$) correlated with each other in the direction that was hypothesized. The level of correlation for Medicaid plans on the spirometry measure and the pharmacotherapy management for COPD exacerbations systemic corticosteroids indicator was moderate (the correlation coefficient was 0.3), while the other correlations were weaker (0.1 or 0.2). Correlations also were weaker for Medicare plans compared to Commercial and Medicaid plans.
- The developer provided updated construct validity results in the current maintenance review:

- For this measure, the developer specifically hypothesized: Performance on the spirometry measure would be positively correlated with performance on the ***Appropriate Testing for Pharyngitis measure***. The second performance measure the developer hypothesized would positively correlate with Spirometry testing for for COPD diagnosis: ***Controlling High Blood Pressure Measure for Medicare Plans***. Health Plan Level Pearson Correlation Coefficients Among Use of Spirometry Testing in Assessment and Diagnosis of COPD and Appropriate Testing for Pharyngitis Measure Performance Scores – Medicaid Plans, 2017*
- Pearson's correlation coefficient between the Spirometry Measure and the Pharyngitis measure for 162 Medicaid plans was reported at 0.48
- Pearson's correlation coefficient between the Spirometry Measure and the Pharyngitis measure for 344 commercial plans was 0.24
- Pearson's correlation coefficient between the Spirometry Measure and the Blood Pressure Measure for 355 Medicare plans was 0.25

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The NQF staff is satisfied with the reliability testing for the measure. Does the Committee agree with the staff recommendation?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff or is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

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

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

Guidance from the Validity Algorithm: 1→2→5→6→7 (eligible rating of HIGH)

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

2a1. Reliability – Specifications

Comments:

** No concerns regarding reliability, but wonder about the rationale/evidence for the time period of 2 years prior to Index episode start date through 6 months after Index episode start date.

** The method of inclusion or exclusion from the numerator and denominator are stable in this maintenance of endorsement submission. There is a track record of consistent implementation in a large number of commercial, Medicare, and Medicaid health plans since its initial implementation in 2006.

** Reliability measures show good reliability across the three types of payers.

** My concerns about this measure being consistently implemented have to do with the end users and not the measure itself.

**This is a pretty simple measure that should be able to be implemented

2a2. Reliability – Testing

Comments:

**No concerns about reliability based on data presented

**Why are individuals with non-acute hospital admissions excluded?

**No

**No

**No concerns

2b1. Validity – Testing

Comments:

**No concerns about validity

**Previous and recent validity tests of the data submitted under HEDIS expectations are provided and appear to be valid.

**The attempt to compare to Appropriate Testing for Pharyngitis seemed low and the issue of the need to scale the white vs. non-white population for the spirometry test seemed to be concerning for validity.

**No

**No concerns

2b4-7. Threats to Validity

Comments:

**No remarkable threats to validity

**The analysis exploring meaningful differences is described on pages 47 & 48. There is a single set of specifications. The HEDIS audit process limits missing data.

**Validity issues were covered appropriately.

**I am not aware of any potential threats to validity.

**The measure is based on claims data, which should minimize messing data

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

**Hospice patients excluded, but this does not pose a remarkable threat to validity

**HEDIS data collection excludes hospice patients. See question #7 above regarding the exclusion of patients with non-acute hospital admissions. Risk adjustment was not employed.

**Not specifically validity issues, but concern about the differences for diferent racial groups and underlying risks due to social risks.

**I agree with the findings of the developer and the reviewers.

**Exclusion of hospice patients seems appropriate

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0577

Measure Title: Use of Spirometry Testing in the Assessment and Diagnosis of COPD

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_xxxx" 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.**

   No concerns. The developer provided updated measure score reliability and construct validity which were calculated from the 347 Commercial health plans, 361 Medicare plans and 173 Medicaid health plans that submitted data on this measure to HEDIS in 2018. The plans were geographically diverse and varied in size.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" 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 <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

The developer provided reliability resting using HEDIS health plan performance data for 2018.

The developer conducted performance measure score reliability testing by using a beta-binomial model (signal to noise ratio)

7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   The developer provided

   - Commercial: the mean reliability result was 0.76 , and ranged 0.48-0.96 in the 10th to 90th percentiles
   - Medicare: the mean reliability result was 0.9, and ranged 0.72-0.99 in the 10th to 90th percentiles
   - Medicaid: the mean reliability result was 0.88 , and ranged 0.66-0.98 in the 10th to 90th percentiles

The developer explained that 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 three indicators within this measure have good reliability between 0.9 and 1.0

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.**

    Methodology and results are appropriate and yielded good reliability results.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Developer excludes hospice patients from the measure, but did not perform an analysis related to exclusions of these patients.

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

Submission document: Testing attachment, section 2b4.

The developer indicates there is variation in performance. To determine if this difference is statistically significant, the developer calculated an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile.The results provided by the developer indicate there is an 8-11% gap in performance between the 25th and 75th percentile performing plans across the different product lines. For most product lines, the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance is in Medicare plans, which shows an 11-percentage point gap between 25th and 75th percentile plans.

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.
N/A

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

No concerns with missing data. In 2018, NCQA auditors did not find any missing data sources for any of the health plan data submissions.

The developer noted HEDIS addresses missing data in a structured way through its audit process. 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.

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**

N/A

VALIDITY: TESTING

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

    ☒ Face validity

    ☒ Empirical validity testing of the measure score

    ☐ N/A (score-level testing not conducted)
19. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    - The developer did updated validity testing from the 344 Commercial health plans, 355 Medicare plans and 162 Medicaid health plans that submitted data on this measure to HEDIS in 2017. The developer conducted performance measure score validity testing via health plan level Pearson correlation analysis between the measure's two components, as well as statin therapy measures.
        - Pearson correlation coefficients the spirometry measure and the pharyngitis measure were 0.48 and 0.24 for Medicaid and Commercial plans, respectively.
        - Pearson correlation coefficient the spirometry measure and the blood pressure measure was 0.25 for Medicare plans.
        - Performance on Pearson correlation coefficients between these indicators and spirometry measures ranged from 0.24 – 0.48.
        - This suggests low to moderate correlations, which was the result hypothesized by the devolper and suggests some comparability in the quality constructs of the measure indicators.
    - The developer previously did face validity by three expert panels in their previous submission of the measure.

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

    Submission document: Testing attachment, section 2b2.3

    - The results of the Pearson correlation test (on HEDIS data of commercial, Medicare and Medicaid health plans) suggests that performance is somewhat correlated. Coefficients were on the low to moderate for most comparisons, which denotes per developer moderate associations.

    - Previous submission face validity results indicate the technical expert panel showed good agreement that the measure as specified accurately.

21. **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)
22. **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)

23. **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.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
Validity methodology and results are appropriate and yielded appropriate validity scores.

ADDITIONAL RECOMMENDATIONS

25. **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 additional 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.

- Per developer, data is collected during the provision of care and coded by someone other than person obtaining original information
- Per developer, all data elements are in defined fields in electronic claims.
- NCQA conducts audits for all HEDIS collection and reporting processes. NCQA conducts an independenty audit of HEDIS process to verify integrity of HEDIS collection and reporting system. NCQA also uses Policy Clarification Support System to generate ongoing feedback from measure users.
- Per developer, NCQA goals align with NQF that noncommercial uses do not require the consent of the measure developer. An example of such would be use by health care physicians in connection with their own practice. However, commercial use of the 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.

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

3. Feasibility

Comments:

**Measure is feasible

**The collection and analysis of data has been stable since the initial implementation in 2006 (see p. 51).

**Feasible as structured.

**I see nothing that would cause me to think therre is anything about this measure that is not completely feasible.

**I don't have concerns about feasibility

Criterion 4: Usability and Use

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

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**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details [Accountability program(s) – details]

#0577 is used in various public reporting and accountability programs:

- NCQA Health Insurance Plan Rating
- NCQA Quality Compass
- NCQA Health Plan Ratings/Report Cards: #0577 is publicly reported on NCQA website annually. The ratings are generated based on performance on HEDIS measures amond other factors. The developer noted that in 2019 performance information from 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
- NCQA Quality compass tool: The Quality Compass tool using performance rankings from public (Medicare and Medicaid) and private insurance across 50 states. Measure #0577 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.

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

- The developer noted that the denominator language was modified to state COPD diagnosis must be on a discharge claim when identifying a denominator event. This revision occurred following feedback generated from multi-stakeholder advisory panels, public commenting and information from the Policy Clarificcation Support System.

Additional Feedback:

- The developer noted that #0577 remains high a priority measure by NCQA, as illustrated by its use in programs such as Health Plan Rating and Quality Compass. States, employers and regional health quality organizations value this measure (and other HEDIS measures) for shining a light on quality. #0577 was first endorsed by NQF in 2009.

***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 [Impact/trends over time/improvement]

Per the developer, m ore Medicaid plans and fewer Medicare plans reported the measure in 2014 compared to 2013 and 2012, which may help explain why the average performance rates did not substantially improve. The developer noted an opportunity to increase performance rates by increasing attention and utilization of this measures.

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 stated, there were no identified unintended consequences for this measure during testing or since implementation.

Potential harms

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

Additional Feedback:

Feedback from previous endorsement cycle indicate that MAP committee recommended the developer explore creating a composite of all COPD measures and then link that composite with the COPD resource use measure.

Questions for the Committee:

- 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

Comments:

**Measure suitable for public reporting, and likely will be increasingly as quality transparency moves forward

**The findings of this established measure are used internally by participating health plans for quality improvement, in publically published comparisons in quality measures between comparable health plans, and in "global" assessments of health care delivery. See pages 51-54.

**Seems adequately reported to the public.

**This measure is publically reported and is used in various public reporting and accountability programs listed by the developer.

**Developer states this is a high priority measure for NCQA and feedback was solicited and incorporated from multiple stakeholders

4b1. Usability – Improvement

Comments:

**No unintended consequences apparent to this reviewer

**A brief discussion on possible improvement, a possible explanation for the lack of improvement in this measure in the years of its collection, and the absence of any recognized harms are discussed on page 53-54.

**NCQA has used this measure in mulitple tools for evaluating quality. That shows the importance of the measure for identifying high-quality, effiicient programs.

**It would be nice to think that the demonstrated results of this measure could be used to achieve higher levels of use. Anything standing in the way of this would be political in nature.

**More routine use of spirometry can guide patient care decisions on treatment as well as provide meaningful information risk for adverse patient outcomes.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0091 : COPD: Spirometry Evaluation
- 0102 : COPD: inhaled bronchodilator therapy
- 2856 : Pharmacotherapy Management of COPD Exacerbation

Harmonization

- During previous submissions for maintenance of endorsement, the standing committee recommended that measures NQF #0577 and #0091 be fully harmonized. In response, NCQA stated that #0091 and #0577 would be reviewed with their respective measure advisory panels at PCPI and NCQA to assess the need for harmonization. In response, during the 2017 submission for endorsement, NCQA highlighted that #0091 was specified at physician level of analysis and used data from medical records or administrative claims whereas #0577 data source is from claims and electronic clinical data and focuses on health plans. In addition, #0091 measures patients aged 18 and older with COPD diagnosis who have spirometry testing performed at least once during the measurement year whereas #0577 measures percentage of patients aged 40 and older with new COPD diagnosis or an exacerbation who receive confirmatory spirometry testing within 6 months of diagnosis. The developer noted data burden as data is collected from different sources by different entities.
- NCQA staff noted that NQF #0102 is a physician level of analysis medication management measure assessing the percentage of patients age 18 years and older with a diagnosis of COPD and who have an FEV1/FVC < 60% and were prescribed an inhaled bronchodilator during the measurement year while NQF #0577 measures spirometry testing within 6 months following new COPD diagnosis.
- Developer noted that NQF #2856 assesses percentage of patients 40 years and older following an ED visit or COPD exacerbation were prescribed or dispensed appropriate medication within 14 days or 30 days of event. The focus of measures NQF #0102 and NQF #2856 is on medication regiment and NQF #0577 is on confirmatory testing, there is no impact on interpretability or added burden of data collection.

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

5. Related and Competing

Comments:

**Measure 0091 is similar -- COPD, spirometry evaluation. ATS is steward for this measure, which has different eligible population of 18 and older.

**Three other present NQF acceptable measures are different in concept and non-competing. See pages 54-55.

**Although different measures that deal with similar issues, the focus on age seems to key to not harmonizing them. Not sure what would be lost if that were done.

**There are what appears to be related measures that are not necessarily in competition with this measure. The standing committee have made previous recommendations in regards to these measures.

**Yes. I think they are harmonized to the extent possible

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/31/2020

- **No NQF Members have submitted support/non-support choices as of this date.**

Developer Submission

Additional evaluations and submission materials attachments…

Brief Measure Information

NQF #: 0577

Corresponding Measures:

De.2. Measure Title: Use of Spirometry Testing in the Assessment and Diagnosis of COPD

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

De.3. Brief Description of Measure: This measure assesses the percentage of patients 40 years of age and older with a new diagnosis of COPD or newly active COPD, who received appropriate spirometry testing to confirm the diagnosis.

1b.1. Developer Rationale: This measure assesses whether patients considered to have a diagnosis of COPD through the presence of symptoms and risk factors (e.g., dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease) had a spirometry assessment to confirm the diagnosis. The improvement in quality envisioned by the use of this measure is to ensure that patients receive spirometry testing to confirm a COPD diagnosis and determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to guide therapy.

S.4. Numerator Statement: The number of patients with at least one claim/encounter for spirometry during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date.

S.6. Denominator Statement: All patients age 42 years or older as of December 31 of the measurement year, who had a new diagnosis of COPD or newly active COPD during the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year.

S.8. Denominator Exclusions: This measure excludes patients who use hospice services, and those with nonacute inpatient stays.

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Dec 04, 2009 **Most Recent Endorsement Date:** Aug 03, 2016

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

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

SPR_Evidence_Form.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)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0577

Measure Title: Use of Spirometry Testing in the Assessment and Diagnosis of COPD

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 11/8/2019

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**

- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ 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: Spirometry testing to confirm COPD diagnosis for adults 40 years and older with a new diagnosis

☐ Appropriate use measure:

☐ 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.

Patient presents with respiratory symptoms, such as dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease >>> Provider gives patient a spirometry test and uses the spirometry test results to confirm the presence of persistent airflow limitation and thus of COPD >>> Provider also uses the spirometry test results to determine the severity of the disease and airflow limitation, the impact on the patient's health status, and the risk of future events, in order to guide therapy >>> Patients receive appropriate therapy to

reduce COPD symptoms, reduce the frequency and severity of exacerbations. Improved outcomes include avoiding urgent care, emergency department visits, and inpatient stays.

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.)

This measure assesses whether patients with a new diagnosis of COPD or newly active COPD were given a spirometry test to confirm the diagnosis. This measure is based on guidelines and evidence that spirometry should be performed to diagnose airflow obstruction in patients with respiratory symptoms in order to make a clinical diagnosis of COPD and determine appropriate therapy.

****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

Source of Systematic Review: • **Title** • **Author** • **Date**	**2015 Submission:** **Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines:** **Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Bethesda (MD): Global Initiative for Chronic Obstructive Lung Disease, 2015. http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html**

• **Citation, including page number** • **URL**	**2019 Submission:** **Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Global Initiative for Chronic Obstructive Lung Disease, 2020. https://goldcopd.org/gold-reports/** **2015 Submission:** **American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:** **Qaseem A, Wilt T, Weinberger S, Hanania N, Criner G, van der Molen T, Marciniuk D, Denberg T, Schünemann H, Wedzicha W, MacDonald R, Shekelle P. Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Annals of Internal Medicine. 2011;155(3):179-191. doi:10.7326/0003-4819-155-3-201108020-00008** **https://annals.org/aim/fullarticle/479627/diagnosis-management-stable-chronic-obstructive-pulmonary-disease-clinical-practice-guideline** **2015 Submission:** **Institute for Clinical Systems Improvement (ICSI):** **Anderson, B., K. Conner, C. Dunn, et al. 2013. Institute for Clinical Systems Improvement. Diagnosis and Management of Chronic Obstructive Pulmonary Disease (COPD).** **https://www.icsi.org/guidelines__more/catalog_guidelines_and_more/catalog_guidelines/catalog_respiratory_guidelines/copd/** **2019 Submission:** **The ICSI Guidelines have been retired. ICSI has endorsed the Veteran's Affairs/Department of Defense (VA/DoD) Clinical Practice Guideline for the Management of Chronic Obstructive Pulmonary Disease.** **Veteran's Affairs and Department of Defense Clinical Practice Guideline for the Management of Chronic Obstructive Pulmonary Defense:** **The Management of Chronic Obstructive Pulmonary Disease Working Group, The Office of Quality, Safety and Value, VA, Washington, DC, & Office of Evidence Based Practice, US Army Medical Command. (2014). Veteran's Affairs /Department of Defense Practice Guideline for the Management of Chronic Obstructive Pulmonary Disease.** **https://www.healthquality.va.gov/guidelines/CD/copd/VADoDCOPDCPG2014.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.	2019 Submission: **Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines (summary of conclusions):** • COPD should be considered in any patient who has dyspnea, chronic cough or sputum production, a history of recurrent lower respiratory tract infections and/or a history of exposure to risk factors for the disease. Spirometry is required to make the diagnosis; the presence of a post-bronchodilator $FEV_1/FVC<.70$ confirms the presence of persistent airflow limitation, and thus, COPD. Spirometry is the most reproducible and objective measurement of airflow limitation. 2015 Submission: **American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:** • "Recommendation 1: ACP, ACCP, ATS, and ERS recommend that spirometry should be obtained to diagnose airflow obstruction in patients with respiratory symptoms (Grade: strong recommendation, moderate-quality evidence)." 2015 Submission: **ICSI Guidelines:** • Page 8: "The diagnosis of COPD should be suspected based on the patient's medical history and physical examination, but spirometry is required to make a diagnosis and confirm the presence of persistent airflow limitation *(Global Initiative for Chronic Obstructive Lung Disease, 2011 [Guideline]).* Spirometry is an established and important method of measuring lung function for the diagnosis and management of patients with COPD. It is recommended for symptomatic patients at risk of COPD, particularly smokers greater than 45 years of age, and for regular follow-up of patients with documented COPD *(Wilt, 2005 [Systematic Review])*. According to the GOLD criteria, COPD is defined as an FEV1/FVC ratio less than 70% after treatment *(Global Initiative for Chronic Obstructive Lung Disease, 2011 [Guideline])*. Large population screening is not recommended." • Page 10: "It is important to distinguish COPD from asthma, because treatment and prognosis differ. Measurement of pre- and post-bronchodilator FEV1 can assist with this differentiation. In asthma, the spirometric abnormality tends to return to normal with bronchodilators, although this distinction between COPD and asthma is not strictly rigid. If the FEV1/FVC ratio improves to > 70% after bronchial dilation, a diagnosis of COPD can be ruled out. Factors commonly used

	to distinguish COPD from asthma include age of onset, smoking history, triggering factors and occupational history." 2019 Submission: **VA/DoD Guidelines** • Page 22: "We recommend that spirometry, demonstrating airflow obstruction (post-bronchodilator forced expiratory volume in one second/forced vital capacity [FEV1/FVC] <70%, with age adjustment for more elderly individuals), be used to confirm all initial diagnoses of chronic obstructive pulmonary disease (COPD) (Strong For)."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	2019 Submission **GOLD Guidelines, 2020** The GOLD report does not assign a grade to its statements on spirometry, nor to the associated evidence. 2015 Submission **American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:** "Evidence is considered moderate quality when it is obtained from RCTs with important limitations—for example, biased assessment of the treatment effect, large loss to follow-up, lack of blinding, unexplained heterogeneity (even if it is generated from rigorous RCTs), indirect evidence originating from similar (but not identical) populations of interest, and RCTs with a very small number of participants or observed events. In addition, evidence from well-designed controlled trials without randomization, well-designed cohort or case– control analytic studies, and multiple time series with or without intervention are in this category. Moderate-quality evidence also means that further research will probably have an important effect on confidence in the estimate of effect and may change the estimate." 2015 Submission **ICSI Guidelines GRADE System Evidence Definitions:** • High Quality Evidence = Further research is very unlikely to change our confidence in the estimate of effect. • Moderate Quality Evidence = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. • Low Quality Evidence = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate or any estimate of effect is very uncertain.

	2019 Submission **VA/DoD Guidelines** Strong For indicates that the Work Group is highly confident that desirable outcomes outweigh undesirable outcomes.
Provide all other grades and definitions from the evidence grading system	2019 Submission **GOLD Guidelines, 2020** The GOLD report does not assign a grade to its statements on spirometry, nor to the associated evidence. 2015 Submission **American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society:** • Evidence is considered high quality when it is obtained from 1 or more well-designed and well-executed randomized, controlled trials (RCTs) that yield consistent and directly applicable results. This also means that further research is very unlikely to change confidence in the estimate of effect. • Evidence obtained from observational studies would typically be rated as low quality because of the risk for bias. Low-quality evidence means that further research is very likely to have an important effect on confidence in the estimate of effect and will probably change the estimate. 2019 Submission **VA/DoD Guidelines** • Weak-For Recommendation: Indicates that the Work Group is less confident of the balance between desirable and undesirable outcomes • Strong-Against Recommendation: Indicates that the Work Group is confident that the undesirable consequences outweigh the desirable consequences • Weak-Against Recommendation: Indicates that the Work Group is less confident that the undesirable consequences outweigh the desirable consequences
Grade assigned to the **recommendation** with definition of the grade	2019 Submission **GOLD Guidelines, 2020** The GOLD report does not assign a grade to its statements on spirometry, nor to the associated evidence.

	2015 Submission **American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society Guidelines:** Strong - Benefits clearly outweigh risks and burden or vice versa. 2015 Submission **ICSI Guidelines GRADE System:** • Guideline • Systematic Review 2019 Submission **VA/DoD Guidelines** The VA/DoD Guidelines do not grade evidence and recommendations separately.
Provide all other grades and definitions from the recommendation grading system	2019 Submission **GOLD Guidelines, 2020** The GOLD report does not assign a grade to its statements on spirometry, nor to the associated evidence. 2015 Submission **American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society Guidelines:** • Weak - Benefits finely balanced with risks and burden. 2019 Submission **VA/DoD Guidelines** The VA/DoD Guidelines do not grade evidence and recommendations separately.
Body of evidence: Quantity – how many studies? Quality – what type of studies?	2015 Submission GOLD Guidelines: 1965 – 2014 2019 Submission **GOLD Guidelines, 2020** The conclusions regarding spirometry from the GOLD guidelines are supported by two observational studies, as well as a summary article from an American Thoracic Society/European Respiratory Society (ATS/ERS) series on interpretative strategies for lung function tests. Although GOLD referenced no

RCTs in their statement supporting use of spirometry, the studies listed are of high quality and present compelling evidence.

2015 Submission

American College of Physicians (ACP), American College of Chest Physicians, American Thoracic Society, and European Respiratory Society: 1966 – 2009

2015 Submission

ICSI Guidelines: 1968-2012

The guideline developers did not provide a breakdown of the specific number of randomized control trials (RCTs) and given the number of studies included in the systematic reviews we were not able to delineate all RCTs for each recommendation. The GOLD guidelines referenced a total of 613 studies to update the previous set of guidelines from 2013. The recommendation for spirometry to confirm a COPD diagnosis was based on a systematic review and three observational studies. The American College of Physicians et al. guidelines referenced a total of 62 studies to update the previous set of guidelines from 2007. The recommendation for spirometry to confirm a COPD diagnosis cited 17 randomized control trials, meta-analyses, systematic reviews and observational studies. The ICSI Guidelines referenced a systematic review related to spirometry testing to confirm a COPD diagnosis.

Although none of the guidelines included a detailed

summary of the quality of the evidence, there is strong evidence from case control and other observational studies that spirometry testing improves diagnostic accuracy of COPD compared to diagnosing COPD based on symptoms alone. The quality of evidence that spirometry testing to confirm a COPD diagnosis is associated with better treatment outcomes is moderate and based on findings from RCTs. One systematic review of three RCTs that enrolled over 2,500 subjects over 3+ years found that pharmacologic treatment effectiveness was associated with disease severity as measured by baseline spirometry; therefore, spirometry testing is useful to identify those patients who might benefit from pharmacologic treatment in order to improve outcomes. The systematic review also found moderate quality evidence from five cohort studies that baseline spirometry testing provides independent prognosis of overall respiratory symptoms and morbidity and mortality in individuals with established COPD and that spirometry results may be useful as a guide for initiation of inhaled medications and pulmonary rehabilitation among individuals having symptoms, especially frequent exacerbations.

2019 Submission

VA/DoD Guidelines

The evidence base to support a recommendation on spirometry included 2 systematic reviews, 1 RCT, 3 case series and 1 cohort trial. Evidence questions guided a systematic evidence review, which identified the body of evidence relevant to each evidence question. The overall quality of the

	body of evidence was assessed using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) methodology, which takes multiple factors (overall study quality, consistency of evidence, directness of evidence, and precision of evidence) into consideration to rate the overall quality of the evidence (i.e. high, moderate, low, very low).
Estimates of benefit and consistency across studies	2015 Submission (across studies) NCQA's measure, Use of Spirometry Testing in the Assessment and Diagnosis of COPD is based on the research literature, guidelines, and expert feedback. Though COPD is a major cause of morbidity and mortality, studies have found that the disease is under-diagnosed, particularly in its milder forms. A number of studies have found that spirometry is a valuable tool for the diagnosis of COPD. One study found that 42% of newly diagnosed cases in study participants would not have been detected without spirometry. Spirometry is particularly useful in distinguishing COPD from asthma. Major clinical guidelines designate spirometry as the gold standard for diagnosis of COPD. On an initial visit for COPD assessment, spirometry assessments can confirm the presence and any reversibility of airflow obstruction (low FEV1 and FEV1/FVC ratio) and distinguish COPD from asthma. There is strong evidence that spirometry testing improves diagnostic accuracy of COPD compared to diagnosing COPD based on symptoms alone; for instance, a systematic review found that a third of patients with normal airflow reported respiratory symptoms and 21 percent with severe airflow obstruction did not report respiratory symptoms. Another benefit of spirometry is to identify those with symptomatic, severe airflow obstruction who might benefit from pharmacologic treatment in order to improve or lessen the number of COPD exacerbations. Compared to diagnosis and treatment based on clinical examination alone, spirometry is likely to reduce the number of individuals reporting symptoms who are inaccurately diagnosed with, and treated for, COPD because they do not have airflow obstruction of severity where treatment is beneficial. The evidence supports the initiation of inhaled bronchodilator treatment (anticholinergics, long-acting β-agonists, or corticosteroids) in patients who have respiratory symptoms and FEV1 less than 60% predicted. 2019 Submission (across studies) Studies have consistently shown that spirometry is appropriate and required for diagnosing COPD.
What harms were identified?	2015 Submission (across studies) The physical performance of spirometry has not been associated with adverse effects. In addition, spirometry testing is relatively inexpensive. The evidence also states that spirometry testing to confirm COPD diagnosis reduces the risk of over-diagnosing and over-treating patients for

	COPD. There is consensus that the benefits of spirometry testing far outweigh the potential harms. 2019 Submission (across studies) Few patients who have asthma rather than COPD may be mis-diagnosed. No new significant risks have been identified, and consensus remains that the benefits of spirometry far outweigh potential harms.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	2019 Submission (across studies) No new studies that change the recommendations stated above have been identified.

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:

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

1b.1. Briefly explain the rationale for this measure *(e.g., 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 patients considered to have a diagnosis of COPD through the presence of symptoms and risk factors (e.g., dyspnea, chronic cough or sputum production, and a history of exposure to risk factors for the disease) had a spirometry assessment to confirm the diagnosis. The improvement in quality envisioned by the use of this measure is to ensure that patients receive spirometry testing to confirm a COPD diagnosis and determine the severity of the disease, its impact on the patient's health status and the risk of future events (such as exacerbations, hospital admissions or death), in order to guide therapy.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for 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. Performance data is summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year and product line (i.e. commercial, Medicaid, and Medicare).

Commercial Rate

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range

2016 | 41.2% | 7.3% | 21.1% | 32.8% | 36.4% | 40.7 % | 44.8% | 49.8% | 72.1% | 8.4%

2017 | 40.5% | 7.1% | 15.2% | 32.2% | 36.3% | 39.9% | 44.1% | 49.5% | 73.6% | 7.8%

2018 | 41.7% | 7.2% | 26.3% | 32.3% | 36.9% | 40.0% | 44.1% | 49.7% | 76.3% | 7.2%

Medicare Rate

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range

2016 | 34.1% | 9.8% | 6.0% | 22.3% | 27.9% | 33.6% | 39.4% | 45.3% | 81.0% | 11.5%

2017 | 34.0% | 10.7% | 3.3% | 21.8% | 28.2% | 33.3% | 40.0% | 45.2% | 76.9% | 11.8%

2018 | 34.9% | 10.7% | 1.3% | 23.0% | 28.8% | 34.3% | 40.0% | 48.1% | 76.4% | 11.2%

Medicaid Rate

YEAR| MEAN | ST DEV | MIN | 10TH | 25TH | 50TH | 75TH | 90TH | MAX | Interquartile Range

2016 | 31.7% | 8.7% | 14.9% | 22.5% | 25.9% | 30.3% | 36.6% | 44.6% | 63.3% | 10.8%

2017 | 31.6% | 9.3% | 10.9% | 20.1% | 25.4% | 30.8% | 36.8% | 42.2% | 65.4% | 11.4%

2018 | 31.1% | 9.1% | 04.6% | 20.7% | 25.5% | 30.4% | 35.1% | 41.1% | 68.1% | 9.6%

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2018, HEDIS measures covered more than 190 million people. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Commercial

YEAR | N Plans | Mean Denominator Size per plan

2016 | 360 | 744

2017 | 347 | 697

2018 | 338 | 707

Medicare

YEAR | N Plans | Mean Denominator Size per plan

2016 | 330 | 992

2017 | 361 | 981

2018 | 242 | 852

Medicaid

YEAR | N Plans | Mean Denominator Size per plan

2016 | 157 | 744

2017 | 173 | 565

2018 | 163 | 1,010

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

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for 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.*

The CMS Office of Minority Health in collaboration with the RAND Corporation produces an annual report: CMS Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage. We provide below summary data for this measure from that report. The authors note that "for reporting HEDIS data stratified by race and ethnicity, racial and ethnic group membership is estimated using a methodology that combines information from CMS administrative data, surname, and residential location."

In the 2019 report, Asian/Pacific Islander patients and Black patients with a new diagnosis of COPD or newly active COPD were more likely than White patients with a similar new or newly active diagnosis to have received a spirometry test to confirm the diagnosis. Hispanic patients were found to be less likely than White patients to receive spirometry.

2019 CMS Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage report. https://www.cms.gov/About-CMS/Agency-Information/OMH/Downloads/2019-National-Level-Results-by-Race-Ethnicity-and-Gender.pdf

HEDIS data are stratified by type of insurance (e.g. commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity, and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities.

Escarce, J.J., Carreon, R., Veselovskiy, G., Lawson, E.G. Collection of Race and Ethnicity Data by Health Plans has Grown Substantially, but Opportunities Remain to Expand Efforts. Health Affairs (Millwood) 2011; 30(10):1984-91. http://www.ncbi.nlm.nih.gov/pubmed/21976343

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

Spirometry itself is often "corrected" for race, by either using a scaling factor for all people not considered to be 'white', or by applying population-specific norms. Spirometer operators must select the race of the patient, either by asking the patient to self-identify, or 'eyeballing' (Braun, 2015). Further, differential access may impact use of the spirometry test. In a review article discussing disparities in COPD, the authors note "…relatively few pulmonologists and specialty services such as spirometry and pulmonary rehabilitation are available to disadvantaged people" (Pleasants, et al, 2015). In a recent cross-sectional study evaluating the

effects of race and gender on the likelihood of a prior diagnosis of COPD, African-Americans had higher odds of not having a prior COPD diagnosis at across all stages of airflow obstruction, compared to non-Hispanic whites. Women had higher odds of having a prior COPD diagnosis at all stages compared to men. The authors concluded that both race and gender are associated with significant disparities in COPD diagnosis (Mamary, et al, 2018).

Braun, L. (2015). Race, ethnicity and lung function: a brief history. Can J Respir Ther, 51(4), 99-101.

Mamary, A. J., Stewart, J. I., Kinney, G. L., Hokanson, J. E., Shenoy, K., Dransfield, M. T., Foreman, M.G., Vance, G.B., Criner, G. J. (2018). Race and gender disparities are evident in COPD underdiagnoses across all severities of measured airflow obstruction. Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation, 5(3), 177-184. doi:10.15326/jcopdf.5.3.2017.0145

Pleasants, R., Riley, I., & Mannino, D. (2016). Defining and targeting health disparities in chronic obstructive pulmonary disease. International Journal of Chronic Obstructive Pulmonary Disease, Volume 11, 2475-2496. doi:10.2147/copd.s79077

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):*

Respiratory, Respiratory : Chronic Obstructive Pulmonary Disease (COPD)

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

Primary Prevention

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

Elderly, Populations at Risk

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 0577_SPR_Value_Sets_Fall_2019.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. <u>For maintenance of endorsement:</u> 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.

No

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

N/A

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

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

The number of patients with at least one claim/encounter for spirometry during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date.

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)*

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

Identify the number of patients with at least one claim/encounter for spirometry (Spirometry Value Set) during the 730 days (2 years) prior to the Index Episode Start Date through 180 days (6 months) after the Index Episode Start Date.

The Index Episode Start Date is the earliest date of service for an eligible visit (outpatient, ED or acute inpatient) during the 6 months prior to the beginning of the measurement year through 6 months after the beginning of the measurement year with any diagnosis of COPD.

- For an outpatient, observation or ED visit, the Index Episode Start Date is the date of service.

- For an acute inpatient encounter identified only by a professional claim (where the discharge date cannot be determined), the Index Episode Start Date is the date of service.

-For an acute inpatient discharge, the Index Episode Start Date is the date of discharge.

-For an acute inpatient discharge with a direct transfer, the Index Episode Start Date is the discharge date of the original admission.

See corresponding Excel file for value set referenced above.

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

All patients age 42 years or older as of December 31 of the measurement year, who had a new diagnosis of COPD or newly active COPD during the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year.

S.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.)*

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

The eligible population for the denominator is defined by following the series of steps below:

Step 1: Determine the Index Episode Start Date. Identify all patients who had any of the following during the intake period (the 6 months prior to the beginning of the measurement year through the 6 months before the end of the measurement year):

1) An outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set), or an ED visit (ED Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include outpatient, ED or observation visits that result in an inpatient stay.

2) An acute inpatient encounter (Acute Inpatient Value Set) with any diagnosis of COPD (COPD value set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set).

3) An acute inpatient discharge with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set) on the discharge claim. To identify acute inpatient discharges:

a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)

b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)

c. Identify the discharge date for the stay.

If the patient had more than one eligible visit, include only the first visit.

Step 2: Test for negative diagnosis history. Exclude patients who had any of the following during the 730-day period prior to the Index Episode Start Date.

1) An outpatient visit (Outpatient Value Set), a telephone visit (Telephone Visits Value Set),and online assessment (Online Assessments Value Set), an observation visit (Observation Value Set), or an ED visit (ED Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set). Do not include outpatient, ED or observation visits that result in an acute inpatient stay.

2) An acute inpatient encounter (Acute Inpatient Value Set) with any diagnosis of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set).

3) An acute inpatient discharge with any diagnoses of COPD (COPD Value Set), emphysema (Emphysema Value Set) or chronic bronchitis (Chronic Bronchitis Value Set) on the discharge claim. To identify acute inpatient discharges:

a. Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set)

b. Exclude nonacute inpatient stays (Nonacute Inpatient Stay Value Set)

c. Identify the discharge date for the stay.

For an acute inpatient discharge Index Episode Start Date, use the Index Episode Start Date of admission to determine the 730-day period. For direct transfers, use the admission date of the original admission to determine the 730 days prior to the Index Episode Start Date.

See corresponding Excel file for value sets referenced above.

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

This measure excludes patients who use hospice services, and those with nonacute inpatient stays.

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 patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began. These patients may be identified using various methods, which may include but are not limited to enrollment data, medical record, claims/encounter data (Hospice Encounter Value Set, Hospice Intervention Value Set).

Exclude patients with nonacute inpatient stays (Nonacute Inpatient Stay Value Set).

See attached Hospice Encounter Value Set, Hospice Intervention Value Set, and Nonacute Inpatient Stay Value Set.

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.)*

N/A

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:

Rate/proportion

If other:

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.)*

The measure calculation is detailed in the steps listed below:

Step 1: Determine the eligible population.

A. Determine the Index Episode Start Date. Identify all patients who had an outpatient visit, observation visit, ED visit, or acute inpatient encounter/discharge with a diagnosis of COPD, emphysema, or chronic bronchitis.

If the patient had more than one eligible visit, include only the first visit.

B. Test for negative diagnosis history.

Step 2: Determine the numerator. Identify the number of patients who had at least one claim/encounter for spirometry.

Step 3: Calculate the rate: Numerator/Denominator

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)*

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

SPR_nqf_testing_attachment_7.1-637088207214677482.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)

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Date of Submission: 8/15/2019

Type of Measure:

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

Instructions

- **Measures must be tested for all the data sources and levels of analyses that are specified. *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* about how to present all the testing information in one form.**
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.***
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**
- **For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.**

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

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

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

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

2b3. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

- **rationale/data support no risk adjustment/ stratification.**

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful [16] differences in performance;

OR

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

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

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

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

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

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

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

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.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*).

N/A

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

Initial testing: During measure development, we conducted a comprehensive field test in 2004 to assess feasibility of data collection and validity of performance data and critical data elements. This field test used data from measurement year 2003, which included health plan data spanning January 1, 2001 through December 31, 2003.

Systematic evaluation of face validity: The measure was tested for face validity throughout measure development from 2004 to 2006. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2006.

Measure score reliability and construct validity testing: We assessed measure score reliability and construct validity using data from all health plans that submitted HEDIS data to NCQA for this measure in 2018, which used data for measurement year 2017.

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*)

Initial testing: To assess feasibility of data collection and validity of performance data and critical data elements, 5 Commercial health plans, 1 Medicaid health plan and 3 Medicare health plans provided individual member-level data to NCQA for analysis. These plans were selected because they had the resources to generate the files, had sufficient sample of members with persistent asthma for analysis, and willingness to provide the data. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity: Throughout the entire measure development process from 2004-2006, the measure was tested for face validity using panels of experts with specific clinical, methodologic and operational expertise. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2006. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliation of expert panel:

1) NCQA's Respiratory Measurement Advisory Panel (RMAP) is comprised of 10 experts (8 physicians, 1 pharmacist and 1 researcher) in clinical pulmonary care, including health care providers and policy makers.

2) NCQA's Technical Measurement Advisory Panel is a 12-member panel representing health plans methodologists, clinicians and HEDIS auditors.

3) NCQA's Committee on Performance Measurement (CPM) oversees the HEDIS measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 17 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

4) NCQA's HEDIS Expert Coding Panel reviewed and provided feedback on the vocabularies and definitions found in the values sets used to identify each measure component as well as the more recent mapping of ICD-9 codes to ICD-10 codes.

In 2005, the draft measure was posted for public comment, a 30-day period of review that allowed interested parties to offer feedback to NCQA about the measure. Stakeholders from various types of organizations submitted 79 comments on the measure.

2019 Update: Measure score reliability and construct validity were calculated from the 347 Commercial health plans, 361 Medicare plans and 173 Medicaid health plans that submitted data on this measure to HEDIS in 2018. The plans were geographically diverse and varied in size.

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

Patient sample for initial measure field testing: We collected data from 5 Commercial health plans, 3 Medicare health plans and 1 Medicaid health plan to assess feasibility of data collection and validity of performance data and critical data elements. Below is a description of the sample. It includes the number of health plans that provided data for the measurement year 2003 and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial	5	157
Medicare	3	414
Medicaid	1	347

2019 Update

Patient sample for measure score reliability and construct validity: Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the number of health plans that submitted data for this measure to HEDIS in 2018 (for measurement year 2017) and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median Number of Eligible Patients per Plan
Commercial	347	697
Medicare	361	566
Medicaid	173	700

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

During measure development, we conducted a comprehensive field test in 2004 to assess feasibility of data collection and validity of performance data and critical data elements using data submitted by 5 Commercial health plans, 1 Medicaid health plan and 3 Medicare health plans. The field test used the measurement year 2003 that included data from January 1, 2001 through December 31, 2003.

Face validity was demonstrated through a systematic assessment of face validity during measure development and at regular intervals since then. Per NQF instructions we have described the composition of the technical expert panel, which assessed face validity in the data sample questions above.

2019 Update

No difference in the data used for reliability and construct validity testing.

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.

2019 and 2015 Submission:

Social risk factor data were not available in reported results. This measure is specified to be reported separately by Medicare, Medicaid and commercial plan types, which serves as a proxy for income and other socioeconomic factors.

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*)

2019 and 2015 Submission:
We utilized the Beta-binomial model (Adams 2009) to assess how well one can confidently distinguish the performance of one accountable 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 in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS measures. 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 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*)

2019 Update:

Reliability statistics for this measure were calculated using HEDIS health plan performance data for 2018. The results are as follows:

Beta-Binomial Statistics:

Commercial			Medicare			Medicaid		
Avg.	Overall	10th-90th	Avg.	Overall	10th-90th	Avg.	Overall	10th-90th
0.76	0.91	0.48-0.96	0.90	0.98	0.72-0.99	0.88	0.95	0.66-0.98

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

Interpretation of measure score reliability testing:

Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, 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 three indicators within this measure have good reliability between 0.9 and 1.0.

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)*

METHOD OF ASSESSING FACE VALIDITY: NCQA tested the measure for face validity using a panel of stakeholders with relevant clinical expertise and research and measurement, experience. This panel included representatives from key stakeholder groups, including the CDC, pulmonologists, provider and delivery organizations and researchers (See list of current members for the Respiratory Advisory Panel (RMAP) under section Ad.1). RMAP experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

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

This measure was developed in 2004 to assess whether patients had spirometry testing to confirm a COPD diagnosis and determine the severity of the disease and its impact on the patient's health status. NCQA and the Respiratory Measurement Advisory Panel worked together to develop the most appropriate measure for assessing spirometry testing to confirm new COPD diagnoses.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.
The spirometry measure was written and field-tested in 2004. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in 2005.

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

The spirometry measure was released for Public Comment in 2005 prior to publication in HEDIS. We received and responded to 79 comments on this measure. The CPM recommended moving this measure to first year data collection by a majority vote.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

The spirometry measure was introduced to HEDIS in 2005. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure to public reporting with a majority vote in 2006.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publicly reported and may be used for scoring in accreditation.
The spirometry measure has been publicly reported in HEDIS since 2006.

STEP 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

The clinical guideline recommendations for spirometry testing in the assessment and diagnosis of COPD have not changed since the measure was developed in 2005; therefore, we have not made any significant changes to the spirometry measure since it was last endorsed on January 31, 2012.

Expert Participation
This measure was tested for face validity with input from three expert panels. Guidelines from the Global Initiative for Chronic Obstructive Lung Disease (GOLD) were also a strong authoritative source in applying the evidence for the spirometry measure.

We list an overview of each panel here. Please refer to Ad.1 in the submission form for the names and affiliation of experts in each panel.

1. Respiratory Measurement Advisory Panel includes 10 members (eight physicians, one pharmacist and a researcher) with expertise in respiratory care and quality measurement.

2. The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.

3. NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA

and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

ICD-10 CONVERSION:
In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process
1. NCQA first identified value sets within the measure that included ICD-9 codes. We used General Equivalence Mapping (GEM)to identify ICD-10 codes that map to ICD-9 codes and reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA then searched for additional codes (not identified by GEM mapping step) that should be considered due to the expansion of concepts in ICD-10. Using ICD-10 tabular list and ICD-10 Index, searches by diagnosis or procedure name were conducted to identify appropriate codes.
3. NCQA HEDIS Expert Coding Panel review: Updated value set recommendations were presented to for expert review and feedback.
4. NCQA RMAP clinical review: Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is consistent and appropriate given the scope of the measure.
5. New value sets containing ICD-10 code recommendations were for public review and comment in 2014 and updated in 2015. Comments received were reconciled with additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
6. NCQA staff finalized value sets containing ICD-10 codes for publication in 2015.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under **Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development**.

METHOD OF TESTING CRITICAL DATA ELEMENT VALIDITY: For the initial field test in 2004, validity was tested by comparing the presence of administrative claims codes for a new diagnosis of COPD (required to calculate the denominator) and for a spirometry test performed (required to calculate the numerator) to documentation in the medical record, which is considered to be the "gold standard".

2019 Update:
METHOD OF TESTING CONSTRUCT VALIDITY: We tested for construct validity by exploring whether this measure was correlated with other similarly constructed process measures. We hypothesized that organizations that perform well on the measure should perform well on other similar HEDIS measures. To test these correlations we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

For this measure, we specifically hypothesized:
1. Performance on the spirometry measure would be positively correlated with performance on the Appropriate Testing for Pharyngitis measure. Examining this correlation would help contribute to the validity of the measure, because both measures assess the receipt of testing to confirm a diagnosis of a respiratory condition. Because the Appropriate Testing for Pharyngitis measure is limited to Commercial and Medicaid plans, we conducted the correlation for those two product lines.

2. Performance on the spirometry measure will be positively correlated with performance on the Controlling Blood Pressure measure for the Medicare product line. Examining this correlation would help contribute to the validity of this measure, because both measures appropriate testing following a chronic disease diagnosis.

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

RESULTS OF FACE VALIDITY ASSESSMENT:

For the initial field test in 2004, we calculated the measure performance rate and discussed the validity of the performance results with the expert panels (Respiratory Measurement Advisory Panel and the Committee on Performance Measurement). Performance rates were between 30 to 32 percent among the different product lines. Performance rates were lower for men compared to women and for people ages 75+ compared to ages 40-74. The expert panels agreed that the performance on the spirometry measure was an accurate representation of quality performance and distinguished performance among health plans.

2004 Field Test: Performance Rates on the Spirometry Measure by Product Line, Age and Gender *

	Denominator	Numerator	Performance Rate
Product Line			
Commercial	3559	1141	32.1%
Medicare	1403	427	30.4%
Medicaid	347	109	31.4%
Age			
40-54	1759	580	33.0%
55-64	1722	563	32.7%
65-74	1029	331	32.2%
75-84	625	177	28.3%
85+	174	26	14.9%
Gender			
F	2826	947	33.5%
M	2483	730	29.4%
Total	5309	1677	31.6%

**Includes data submitted by 5 Commercial plans, 3 Medicare plans and 1 Medicaid plan using measurement year 2003*

RESULTS OF CRITICAL DATA ELEMENT VALIDTY

Across four plans, validation of a new COPD diagnosis in the medical record was 64%, with a range of 30% to 100%. Validation was higher in the Medicare population (74%) compared to the Commercial population (60%).

2004 Field Test: COPD Diagnosis Medical Record Validation by Plan and Product Line*

	Number of patients with a new COPD Dx using administrative data	% of patients that had documentation of a new COPD Dx in medical record	% of patients that did not have documentation of a new COPD Dx in medical record
Plan:			
A	72	30.0%	48.6%
B	19	68.8%	26.3%
C	20	30.0%	70.0%
D	66	100.0%	0.0%
Total	177	63.8%	31.1%

Product Line:			
Commercial	126	60.2%	32.5%
Medicare	51	73.5%	25.5%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

In four plans, there was 76.6% data consistency for spirometry testing between administrative and medical record data. This was calculated by adding the percent of spirometry tests found in administrative data and medical record data plus the percent of spirometry tests found in neither data source. Contrary to the perceived notion that many spirometry tests happen in the physician office without a claim generated, in two plans no spirometry tests were found in medical records only without a corresponding administrative claim and in the remaining two plans relatively few spirometry tests were found only in medical record data, indicating administrative data are reliable for capturing spirometry tests.

2004 Field Test: Spirometry Numerator Validation by Plan*

Plan Code	# of patients with a new COPD Dx confirmed in both admin & medical record data	% of patients with spirometry confirmed in both medical record & admin data	% of patients with spirometry confirmed in neither admin or medical record data	% of patients with spirometry confirmed in admin data only	% of patients with spirometry confirmed in medical record data only
A	15	46.7%	20.0%	33.3%	0.0%
B	11	54.5%	36.4%	9.1%	0.0%
C	5	50.0%	33.3%	0.0%	20.0%
D	66	18.2%	57.6%	4.5%	19.7%
Total	98	28.6%	48.0%	9.2%	14.3%

**Includes data submitted by 5 Commercial plans and 3 Medicare plans using measurement year 2003*

2019 Update:

STATISTICAL RESULTS OF CONSTRUCT VALIDITY

Health Plan Level Pearson Correlation Coefficients Among Use of Spirometry Testing in Assessment and Diagnosis of COPD and Appropriate Testing for Pharyngitis Measure Performance Scores – **Medicaid** Plans, 2017*

	Spirometry Measure
Pharyngitis Measure	0.48

**Includes data submitted by 162 Medicaid plans using measurement year 2017*

All scores were significant at $p<0.05$

Health Plan Level Pearson Correlation Coefficients Among Use of Spirometry Testing in Assessment and Diagnosis of COPD and Appropriate Testing for Pharyngitis Measure Performance Scores – **Commercial** Plans, 2017*

	Spirometry Measure
Pharyngitis Measure	0.24

**Includes data submitted by 344 Commercial plans using measurement year 2017*

All scores were significant at $p<0.05$

Health Plan Level Pearson Correlation Coefficients Among Use of Spirometry Testing in Assessment and Controlling High Blood Pressure Measure Performance Scores – **Medicare** Plans, 2017*

	Spirometry Measure
Blood Pressure Measure	0.25

**Includes data submitted by 355 Medicare plans using measurement year 2017*
All scores were significant at p<0.05

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?*)

SYSTEMATIC ASSESSMENT OF FACE VALIDITY: The use of spirometry in the assessment and diagnosis of COPD measure was deemed to have the desirable attributes of a HEDIS measure in 2005 (relevance, scientific soundness, and feasibility). The technical expert panels showed good agreement that the measure as specified accurately differentiates quality across providers. The technical expert panels have also reviewed the measure for relevance, scientific soundness, and feasibility at regular intervals since 2005.Our interpretation is that this measure has sufficient face validity.

CRITICAL DATA ELEMENT VALIDTY: The results of the critical data element validity testing demonstrate that the administrative data elements used to calculate the measure denominator (patients with a new diagnosis of COPD) and numerator (patients that had a spirometry test performed) had moderate to strong agreement with medical record data and are valid.

2019 Update

CONSTRUCT VALIDITY: Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. For Medicaid plans, the results confirmed the hypothesis that plans who provide appropriate testing to confirm a diagnosis COPD are more likely to provide appropriate testing to confirm a diagnosis of pharyngitis, suggesting they represent the same underlying quality construct of respiratory quality of care. For commercial plans, the results were slightly less than 0.3 but still showed a positive correlation between the spirometry and pharyngitis measures. For Medicare plans, the results were slightly less than 0.3 but still showed a positive correlation between the spirometry measure and controlling high blood pressure measure, suggesting they may represent the same underlying quality construct of chronic disease quality of care.

2b2. EXCLUSIONS ANALYSIS

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

NCQA has a policy for excluding hospice patients from HEDIS measures.

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*)

N/A

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*)
N/A

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*)

_____

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.

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)*

2019 and 2015 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 standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the

75th percentile of performance. We used these two plans as examples of measured entities. However, the method can be used for comparison of any two measured entities.

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*)

HEDIS 2018 Variation in Performance across Health Plans (Measurement Year 2017)

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	697	40.5	7.1	32.2	36.3	39.9	44.1	49.5	7.8	<0.001
Medicare	982	34.0	10.7	21.9	28.2	33.3	40.0	45.2	11.8	<0.001
Medicaid	566	31.6	9.3	20.1	25.4	30.8	36.8	42.2	11.4	<.0002

EP: Eligible Population, the average denominator size across all plans submitting 2018 HEDIS data for this measure

IQR: Interquartile range

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

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?*)

The results above indicate there is an 8-11% gap in performance between the 25th and 75th percentile performing plans across the different product lines. For most product lines, the difference between the 25th and 75th percentile performance rates is statistically significant. The highest variation in performance is in Medicare plans, which shows an 11-percentage point gap between 25th and 75th percentile plans.

To put these meaningful differences in performance into context, we estimated that on average 115 additional members per Medicare HMO plan would have had a spirometry test performed to confirm COPD diagnosis if plans in the 25th percentile performed as well as plans in the 75th percentile. This estimate is based on the average health plan eligible population.

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

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

Note: *This item is directed to measures that are risk-adjusted (with or without 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*)

N/A

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*)

N/A

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*)

N/A

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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

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

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*)

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.

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 nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

All of the commercial and Medicaid health plans that reported 2018 HEDIS data for this measure reported valid rates as determined by NCQA-certified auditors through the process described above. This means that auditors did not find any missing data sources for any of the health plan data submissions and determined that none of the rates

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) 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 are 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)
	Public Reporting NCQA Health Insurance Plan Rating https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ NCQA Health Insurance Plan Rating https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Quality Improvement (external benchmarking to organizations) NCQA Quality Compass 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 RATINGS/REPORT CARDS: This measure is used in the calculation of health plan ratings, which are reported on the NCQA website annually. These ratings are based on performance on HEDIS measures among other factors.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 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 Health Care Quality Congress), 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 using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, 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.

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

Questions received through the Policy Clarification Support system have generally centered on clarification on whether certain notation in medical record documentation is sufficient to meet numerator criteria. Other questions have sought clarification about continuous enrollment criteria, allowable gaps, and inpatient discharges.

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

This measure has been deemed a priority measure by NCQA, as illustrated by its use in programs such as Health Plan Rating and Quality Compass. States, employers and regional health quality organizations value this measure (and other HEDIS measures) for shining a light on quality.

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.

Feedback obtained through the mechanisms described in 4a2.2.1 informed how we revised the measure for clarity, such as adding language to clarify that a COPD diagnosis must be on a discharge claim when identifying a denominator event.

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.

More Medicaid plans and fewer Medicare plans reported the measure in 2014 compared to 2013 and 2012, which may help explain why the average performance rates did not substantially improve. There is hope that with increasing attention to this measure in public reporting programs, performance rates will improve.

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 unexpected 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)

0091 : COPD: Spirometry Evaluation

0102 : COPD: inhaled bronchodilator therapy

2856 : Pharmacotherapy Management of COPD Exacerbation

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

Pharmacotherapy Management of COPD Exacerbation (National Committee for Quality Assurance)

5a. Harmonization 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.

Our current measure, NQF 0577: Use of Spirometry Testing in the Assessment and Diagnosis of COPD, assesses the percentage of patients ages 40 and older with a new diagnosis of COPD or newly active COPD who received appropriate spirometry testing to confirm the diagnosis. It is a health-plan level measure that uses administrative claims and electronic clinical data from the ambulatory care setting. The following is a description of the differences and the impact on interpretability and data collection burden between our proposed measure and each related measure listed in 5.1a: 0091: COPD: Spirometry Evaluation NQF 0091 assesses the percentage of patients age 18 years and older with a diagnosis of COPD who had spirometry results documented. It is a physician-level measure that uses administrative claims or medical record data. There is some added burden of data collection because the data for each measure is collected from different data sources by different entities. Additionally, the focus of the measures is different. NQF 0091 focuses on

whether patients with a COPD diagnosis (not specifically a new diagnosis) had spirometry testing performed at least once during the measurement year, while NQF 0577 specifies that patients with a new/newly active COPD diagnosis receive spirometry testing to confirm diagnosis. NQF 0102: COPD: Inhaled Bronchodilator Therapy and NQF 2856: Pharmacotherapy Management for COPD Exacerbation Measures NQF 0102 assesses the percentage of patients age 18 years and older with a diagnosis of COPD and who have an FEV1/FVC < 60% and who have symptoms who were prescribed an inhaled bronchodilator. NQF 0102 is a physician-level measure. The NQF 2856 measure assesses the percentage of COPD exacerbations for patients 40 years of age and older who had an acute inpatient discharge or ED encounter during the measurement year and who were dispensed appropriate medications. Two rates are reported. 1. Dispensed a systemic corticosteroid (or there was evidence of an active prescription) within 14 days of the event 2. Dispensed a bronchodilator (or there was evidence of an active prescription) within 30 days of the event Both of these measures focus on medication management for stable COPD or following an exacerbation, while NQF 0577 focuses on appropriate spirometry testing to confirm a new COPD diagnosis. There is no impact on interpretability of publicly-reported rates or added burden of data collection because the focus of our measure is different.

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: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

RESPIRATORY MEASUREMENT ADVISORY PANEL (RMAP) MEMBERS:

David Au, MD, MS, (CHAIR) Associate Prof. of Medicine, VA Puget Sound Health Care System

Kurt Elward, MD, MPH, Senior Medical Director, Innovation Health

Laura Feemster, MD, MS, Investigator/Staff Physician, University of Washington Medical Center

Anne Fuhlbrigge, MD, MS, Senior Associate Dean for Clinical Affairs, University of Colorado School of Medicine

Min Joo, MD, MPH, FCCP, Assistant Professor of Medicine, University of Illinois at Chicago/ Jesse Brown VA Medical Center

Christine Joseph, PhD, MPH, BSc, Associate Director of Research, Epidemiologist, Henry Ford Health System

Todd Lee, PharmD, PhD, Primary: Senior Investigator, Secondary: Associate Professor, University of Illinois at Chicago

Allan Luskin, MD, President, Healthy Airways

Richard O´Connor, MD, Director, Dept. of Quality Management, Allergist/Immunologist, Sharp Rees-Stealy Medical Group

COMMITTEE ON PERFORMANCE MEASUREMENT (CPM) MEMBERS:

Andrew Baskin, MD, National Medical Director, Aetna

Elizabeth Drye, MD, SM, Senior Director, Yale/CORE; Research Scientist, Pediatrics, Yale University

Andrea Gelzer, MD, MS, FACP, Senior VP, Medical Affairs, AmeriHealth Caritas

Kate Goodrich, MD, MHS, Chief Medical Officer and Director, CCSQ, CMS

David Grossman, MD, MPH, Senior Associate Medical Director, Washington Permanente Medical Group

Christine S. Hunter, MD (Co- Chair), Independent Board Director, WPS Health Solutions

David K. Kelley, MD, MPA, Chief Medical Officer, Office of Medical Assistance Programs, Pennsylvania Department of Human Services

Jeffery Kelman, MMSc, MD, Chief Medical Officer, United States Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Freelancer

Adrienne Mims, MD, MPH, AGSF, FAAFP, VP, Chief Medical Officer, Alliant Health Solutions

Amanda Parsons, MD, MBA, Deputy Chief Medical Officer, Metroplus

Wayne Rawlins, MD, MBA, Chief Medical Officer, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Associate Vice President, Clinical Quality Officer, Humana

Rudy Saenz, MD, MMM, FACOG, Physician, Medical Director of Quality, Riverside Medical Clinic

Marcus Thygeson, MD, MPH (Co-Chair), Chief Health Officer, Bind On-Demand

JoAnn Volk, MA, Research Professor, Georgetown University

TECHNICAL MEASUREMENT ADVISORY PANEL MEMBERS:

Andy Amster, MSPH, Senior Director, Kaiser Permanente

Jennifer Brudnicki, MBA, Product Services Manager, Inovalon

Lindsay Cogan, PhD, MS, Director, Division of Quality Measurement, New York State Department of Health

Kathryn Coltin, MPH, Independent Consultant

Michael Farina, R.Ph, MBA, Director, Health Care Quality, Capital District Physicians' Health Plan

Marissa Finn, MBA, CIGNA

Scott Fox, MS, MEd, Principal, Payment Reform, The MITRE Corporation

Carlos Hernandez, Director, Quality Management, CenCalHealth

Harmon Jordan, ScD, Senior Study Director, Westat

Gigi Raney, LCSW, Center for Medicaid and CHIP Services

Lynne Rothney-Kozlak, MPH, President, Rothney-Kozlak Consulting, LLC

Laurie Spoll, Director, Aetna

The NCQA Respiratory Measurement Advisory Panel 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, including a 30-day public comment period, 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: 2005

Ad.3 Month and Year of most recent revision: 07, 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, 2020

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion 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.