NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through National Quality Forum's (NQF) Consensus Development Process (CDP). The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Preliminary Analysis* sections.

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

Brief Measure Information

NQF #: 0091

Measure Title: COPD: Spirometry Evaluation

Measure Steward: American Thoracic Society

Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of COPD who had spirometry results documented.

Developer Rationale:

Current Submission: *The following rationale contains updated information from our recent literature search-:*

"Despite major efforts to broadly disseminate the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines and use of COPD performance measures across different specialty societies, COPD remains underdiagnosed and misdiagnosed (Collins et al., 2015; Perez et al., 2011; Diab et al., 2018; Gershon et al,, 2018; Tisi et.al 2022; Farooqi et al, 2022). Although spirometry use has increased, it remains underutilized to confirm airflow obstruction and accurately diagnose COPD (CDC, 2012; Nishi et al., 2013; Rodwin et al., 2022). Studies show proper COPD diagnosis with spirometry is done on just over half of patients in the US and Canada (Boulet et al., 2013; Bourbeau et al., 2008; Collins et al., 2015; Nishi et al., 2013; Perez et al., 2011; Yu et al., 2013) and ranges from 10-48% in the Asia-Pacific region, Africa, eastern Europe, and Latin America (Aisanov et al., 2012). World-wide, as many of 70% of patients with COPD may be underdiagnosed, while 30-60% of patients are over-diagnosed (Diab et. al, 2018) A study of physician-diagnosed COPD patients hospitalized for exacerbations found that 22% of patients did not have COPD upon spirometry testing (Prieto Centurion, et al., 2012). Treatment of COPD without accurate diagnosis and understanding of true etiology of symptoms results in patients not receiving medication that would improve symptoms and quality of life, prevent exacerbations and reduce costly use of emergency and hospital services. Patients may be exposed to adverse effects of unneeded medication and or delays in true diagnosis and management of another condition increasing overall cost of care (Boulet et al., 2013; Bourbeau et al., 2008; CDC, 2012; Collins et al., 2015; Joo et al., 2011). Several recent studies emphasize the association between both under- and over- diagnosis of COPD with increased respiratory symptoms and health care utilization (Gershon et al, 2018; Farooqi et al, 2022). We believe this measure will continue to increase appropriate spirometry use to assist physicians in the accurate

diagnosis and treatment of patients with COPD, improving patient management and reducing total costs of COPD."

Additional New Citations:

1. Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet LP, and Aaron SD. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018 Nov 1;198(9):1130-1139.
2. Farooqi MAM, Ma J, Ali MU, et al. Prevalence and burden of COPD misclassification in the Canadian Longitudinal Study on Aging (CLSA). BMJ Open Resp Res 2022;9(1):e001156. doi:10.1136/ bmjresp-2021-001156
3. Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, Tan W, To T for the Canadian Respiratory Research Network. Health services burden of undiagnosed and overdiagnosed COPD. Chest 2018; 153(6):1336-1346.
4. Rodwin, BA, DeRycke, EC, Han, L. et al. Characteristics Associated with Spirometry Guideline Adherence in VA Patients Hospitalized with Chronic Obstructive Pulmonary Disease. J Gen Intern Med. 14 Oct 2022.
5. Tisi S, Dickson JL, Horst C, et al. Detection of COPD in the SUMMIT Study Lung Cancer Screening Cohort using Symptoms and Spirometry. Eur Respir J. 2022 Jul 26; online ahead of print (https://doi.org/10.1183/13993003.00795-2022).

Prior Citations:

1. Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.
2. Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.
3. Bourbeau J, Sebaldt RJ, Day A, Bouchard J, Kaplan A, Hernandez P, Rouleau M, et al. Practice patterns in the management of chronic obstructive pulmonary disease in primary practice: the CAGE study. Can Respir J. 2008 JanFeb:15(1):13-9.
4. Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated healthcare resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.
5. Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.
6. Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.
7. Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.
8. Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.
9. Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.
10. Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

Previous 2015 Submission

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

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

Citations:

1. Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.
2. Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.
3. Bourbeau J, Sebaldt RJ, Day A, Bouchard J, Kaplan A, Hernandez P, Rouleau M, et al. Practice patterns in the management of chronic obstructive pulmonary disease in primary practice: the CAGE study. Can Respir J. 2008 Jan-Feb:15(1):13-9.
4. Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.
5. Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.
6. Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.
7. Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.
8. Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.
9. Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.
10. Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

Numerator Statement: Patients with documented spirometry results in the medical record (FEV1 and FEV1/FVC)

Denominator Statement: All patients aged 18 years and older with a diagnosis of COPD

Denominator Exclusions: Documentation of medical reason(s) for not documenting and reviewing spirometry results
Documentation of patient reason(s) for not documenting and reviewing spirometry results
Documentation of system reason(s) for not documenting and reviewing spirometry results

Measure Type: Process

Data Source:

Claims

Level of Analysis:

Clinician: Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: 2009-08-10 12:00 AM

Most Recent Endorsement Date: 8/3/2016 5:09:28 PM

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the measure 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 a change in evidence since the prior evaluation

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

The developer provides the following description for this measure:

- This is a maintenance process measure at the group/practice level that measures the percentage of patients aged 18 years and older with a diagnosis of COPD who had spirometry results documented. The developer provides a logic model statement that says the use of spirometry is important to

confirm the correct diagnosis of COPD. The correct diagnosis leads to appropriate treatment choices which improve patient outcomes, decrease symptoms, reduced exacerbation, and improve health related quality of life.

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- The 2016 Pulmonary and Critical Care Committee agreed with the developer that the underlying evidence for the measure hasn't changed since the last NQF endorsement review in 2012, which included recommendations from the 2011 Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. The Committee agreed the guidelines are clear about using spirometry to confirm the diagnosis of COPD, and not for general screening or monitoring of treatment.
- The Global Initiative for Chronic Obstructive Lung Disease, (GOLD), concurred with similar recommendations to the joint society guidance.
- In 2016 the Pulmonary and Critical Care Committee accepted the prior evaluation rating for evidence.

Changes to evidence from last review

☒ The developer provided updated evidence for this measure:

- The developer summarizes a few studies that found continued underuse of spirometry for confirmation of COPD and highlighted the continued patterns of both under and over-diagnosis of COPD.
- The Global Initiative for Chronic Obstructive Lung Disease GOLD guidelines were updated in 2022 to include that spirometry is now required (instead of support) to make a confident diagnosis of COPD.
- The developer further states that a study published in 2022 regarding the utility of using spirometry to screen asymptomatic individuals for COPD and in 2022 the US Preventive Task Force and the and continue to recommend against the use of routine screening for COPD and therefore NQF #0091 maintains it specifications to confirm the diagnosis of COPD and not as a screening tool.

Exception to evidence

- N/A

Questions for the Standing Committee:

- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Standing Committee agree that there is no need for repeated discussion and a vote on evidence?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance From the Evidence Algorithm

Process measure based on systematic review (SR) (Box 3). A summary of the quantity, quality, and consistency of the body of evidence from the SR is provided in the submission form (Box 4). The SR rated the evidence as moderate (Box 5b). The highest possible rating is high.

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

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

Maintenance measures – increased emphasis on gap and variation

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

- For NQF maintenance of endorsement, measure stewards/developers are expected to provide current performance data on the measure as specified. Data from the literature can be considered if current data are not available.
- The developer provided performance data available for measure NQF #0577, noting that although there are important differences in the measures, the two are similar enough that an identified gap in performance could likely apply to both. Additionally, the developer provided literature to further support that an opportunity for improvement exists.
- Performance data provided for NQF #0577 includes aggregate rates by health plan type (Commercial PPO, Commercial HMO, Medicare HMO etc.) of adults 40 and older who have a new COPD, or newly active COPD, who received spirometry testing to confirm the diagnosis. The data show a decrease in performance in 2020 (pandemic) data compared 2018 for all plan types reported. In both 2018 and 2020, the Medicaid Health Maintenance Organizations (HMO) had the lowest average spirometry testing rates with 31 percent and 26.8 percent respectfully. In 2018 and 2020 Commercial HMO plans had the highest rates reported with averages rates of 41.7 percent and 37.3 percent respectfully. 2018 data is the comparison year since not all plan types reported in 2019.
- The developer also references a study published in 2018 that found worldwide, "up to 70% of patients with COPD may be underdiagnosed, while 30-60 percent of patients are over diagnosed with COPD.

Disparities

- The developer provided a summary of literature that indicates gender and race disparities exist in the diagnosis of COPD, which may be linked to a disparity in performing spirometry on at risk populations.

Questions for the Standing Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**

Measure evaluated by the Technical Expert Panel (TEP)? ☐ **Yes** ☒ **No**

2a. Reliability: Specifications and Testing

For maintenance measures—no change in emphasis—specifications should be evaluated the same as with new measures.

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

For maintenance measures – less emphasis if no new testing data are provided.

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

Specifications:

- Have the measure specifications changed since the last review? ☐ Yes ☒ No
- Measure specifications are clear and precise.

Reliability Testing:

- Did the developer conduct new reliability testing? ☐ Yes ☒ No
- The 2016 Pulmonary and Critical Care Committee passed the measure on reliability with a moderate rating.
- The 2016 Committee expressed concern that the time window indicates a one-year measurement period, but it appears that a spirometry test at any time from age 18 and up counts in the numerator. The developer clarified the goal of the measure is to capture whether the spirometry test was conducted before treatment occurred. The physicians conducting treatment do not necessarily have to perform the test within that year, but need to verify that the test was completed and annually record the results.
- Reliability testing conducted at the accountable entity level:
    - The method used for reliability testing was not described in the submission. The developer should be prepared to speak to the methodology should the Standing Committee choose to discuss and revote on reliability.
    - The developer did not provide updated testing data, using CY 2012 claims data that were used in the 2015 submission. This included over 11.5M Medicare beneficiaries from 2,064 groups of physicians with at least 25 eligible providers (EPs). Groups were included if they had at least 20 eligible cases for the measure.
    - Reliability was tested as a ratio of performance between groups over the total variation. Detailed references and explanations were not provided for the methodology.
    - Reliability for groups with 25 or more EPs (average of 2,974 beneficiaries) was 0.73.
    - Reliability for groups with 100 or more EPs (average of 7,842 beneficiaries) was 0.83.

Questions for the Standing Committee regarding reliability:

- *The developer attests that the specifications have not changed and that additional reliability testing was not conducted. Does the Standing Committee agree that the measure is still reliable and that there is no need for repeated discussion and a vote on reliability?*

Guidance From the Reliability Algorithm

The developer provided complete specifications that can be consistently implemented by users (Box 1) > The empirical reliability was tested using statistical tests with the measure as specified (Box 2) > The empirical reliability testing was conducted at the accountable entity level for each level of analysis (Box 4) > The developer conducted testing method reliability that estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error) (Box 5): There is moderate confidence that the accountable entity levels are reliable (Box 6b) > Moderate. The highest possible rating is high.

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

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

For maintenance measures – less emphasis if no new testing data are provided

2b1. Measure Intent: The measure specifications are consistent with the measure's intent and capture the most inclusive target population.

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.

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

Validity Testing

- Did the developer conduct new validity testing? ☐ Yes ☒ No
- The 2016 Pulmonary and Critical Care Committee passed the measure on validity with a moderate rating.

- Validity testing conducted at the accountable-entity level:
    - The measure was demonstrated using face validity and data element validity. The developer did not provide updated results from the 2015 submission.
    - 12 members of the ATS Clinical Practice Committee rated their agreement if the measure provides an accurate reflection of quality and can be used to distinguish between good and poor quality. Mean rating score was 4.6, with 8= strongly agree, 3 agree, 1 neither agree/disagree (N=12 respondents).
    - The developer also tested the data element generated from the EHR against manually calculated scores by trained abstractors. The Kappa statistic for the numerator (N=123) was 0.7281 (0.60860, 0.8476 CI). They could not calculate a Kappa for the denominator (N=123) because they had 100% agreement. The data have not been updated since the 2015 submission. Note that the measure is specified using claims data and not EHR data.

Exclusions

- The following exclusions are applied to the measure:
    - Documentation of medical, patient or system reason for not documenting and reviewing spirometry results.
    - Of the 123 records reviewed, there was only 1 exclusion (0.81%) and on further review this was found not to be a valid exclusion.
    - Testing of exclusions was not provided due to the lack of any appropriate exclusions.

Risk Adjustment

- The measure is not risk-adjusted or stratified.

Meaningful Differences

- The developer calculated benchmarks using the weighted average of groups with 25 or more EPs and groups with 100 or more EPs.
- 45.6% of groups differed from the benchmark for groups with 25 or more EPs ($p<.05$).
- 47.1% of groups differed from the benchmark for groups with 100 or more EPs ($p<.05$).
- Mean performance for this measure was 54.3%. 25th percentile performance was 17.39% and 75th percentile performance was 83.33%. The interquartile range is 65.94 percentage points.

Missing Data

- The developer states that missing data analysis was not performed for this measure. It is unclear from the developer's response if there is missing data. The developer should be prepared to clarify if there is missing data.

Comparability

- The measure only uses one set of specifications for this measure.

Questions for the Standing Committee regarding validity:

- *The developer attests that additional validity testing was not conducted. Is the sample size (123 records from one provider group) adequate for generalizability?*
- *The developer did not provide validity testing on missing data. The Standing Committee should consider asking the developer to provide a rationale for this. Does the Standing Committee have concerns about the lack of missing data information and testing?*
- *Does the Standing Committee agree that the measure is still valid and that there is no need for repeated discussion and a vote on validity?*

Guidance From the Validity Algorithm

Threats to validity including missing data were not empirically assessed (Box 1).

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

All threats to validity relevant to the measure, including miss data, must be empirically assessed.

Criterion 3. Feasibility

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

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

- The measure is claims based and all data elements are in defined electronic fields. The data is collected by and used by healthcare personnel during the provision of care.

Questions for the Standing 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)?*

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

Criterion 4: Use and Usability

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 evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Current uses of the measure

Publicly reported?	☐ Yes	☒ No			
Current use in an accountability program?	☐ Yes	☒ No	☐ UNCLEAR		
Planned use in an accountability program?	☒ Yes	☐ No	☐ NA		

Accountability program details

- The measure is not currently publicly reported and is not used in an accountability program which are both required for maintenance measures following initial endorsement.
- The measure was included in the Medicare PQRS Program ending in 2016 and more recently used in the Centers for Medicare & Medicaid, (CMS), Merit Based Incentive Program (MIPS) though performance year 2019. CMS removed the measure from MIPS starting in performance year 2020 as documentation of spirometry is a required component of another measure, NQF #102 Appropriate Use of Long-Acting Bronchodilators.
- The developer is scheduling a meeting with CMS to discuss future use in the CMS MIPS Value Pathways related to COPD, asthma, sleep and general pulmonary.

- **4a.2. Feedback on the measure provided 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; and (3) This feedback has been considered when changes are incorporated into the measure.

Feedback on the measure provided by those being measured or others

- CMS publicly reports Quality Payment Program (QPP) performance data for participating healthcare providers (and groups, etc.) in the Provider Data Catalog (PDC). The data was previously available from Physician Compare. CMS included the spirometry measure for Performance years 2017- 2019 and archive data is available.
- The developer has not obtained any specific feedback about the measure from those being measured and other data users and states that CMS did not report problems.
- The developer did not report any feedback from those using the measure and have not made updates to the measure.

***Questions for the Standing 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**

RATIONALE: NQF requires that 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).

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and 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

- Data to support progress on improvement or trends in performance results were not provided.
- The developer reports that the use of spirometry to confirm COPD diagnosis remains low, especially as there is evidence of misdiagnosis (both over and under) of COPD which leads to poor treatment decisions resulting poorer health outcomes.

4b2. Benefits versus harms. 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).

Unexpected findings (positive or negative) during implementation

- The developer states they are not aware of any intended consequences or benefits related to this measure.

Potential harms

- The developer states they are not aware of any intended consequences or benefits related to this measure.

***Questions for the Standing 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: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Rationale: Data to support progress on improvement or trends in performance results were not provided.

Criterion 5: Related and Competing Measures

Related Measures

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

Harmonization

The 2016 Committee felt measure #0091 and #0577 were related and should be harmonized. The Committee agreed that because the measures have similar goals, the developers should consider harmonizing the age limit and timeframe for both the diagnosis of COPD and the timeframe in which the spirometry evaluation is

completed. If recommended for endorsement, the current Committee may be asked to provide additional recommendations for harmonizing the measures.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

---

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]

Yes

[Yes Please Explain]

Since the most recent measure evaluation, several studies have been published that demonstrate continued underuse of spirometry for confirmation of COPD and highlight the continued high prevalence of both under- and over-diagnosis of COPD (Rodwin et al., 2022; Diab N et al., 2018; Gershon AS et al, 2018; Tisi et al., 2022; Farooqi MA et al, 2022) .

There has been additional study into the utility of using spirometry to screen asymptomatic individuals for COPD (Bhatt et al., 2022), but updated US Preventative Task Force recommendations and current guidelines/statements continue to recommend against the use of routine screening for COPD (USPTF 2022; GOLD 2022; Qaseem et al 2011). As a result, our measure remains focused on the use of spirometry for the confirmation of the diagnosis of COPD, rather than as a routine screening tool.

Please also see 1a.12 for more complete detail regarding this new evidence.

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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.

[Response Begins]

Current Submission: The measure's focus is on ensuring that spirometry is used to confirm the diagnosis of COPD (process measure). Receipt of confirmatory spirometry is a necessary step in ensuring that patients with COPD receive

appropriate treatments known to improve patient outcomes, including decreasing symptoms, reducing exacerbations, and improving health-related quality of life.

Our current response is an edited version of the prior submission response, copied here:

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

[Response Ends]

1a.02. Select the type of source for 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.

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

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

https://www.thoracic.org/statements/resources/copd/179full.pdf

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

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

"COPD should be considered in any patient with dyspnea, chronic cough or sputum production, and/or a history of exposure to risk factors for the disease. Spirometry is required to make the diagnosis in this clinical context; the presence of post-bronchodilator FEV1/FVC <0.70 confirms the presence of persistent airflow limitation and thus of COPD in patients with appropriate symptoms and significant exposure to noxious stimuli. Spirometry is the most reproducible and objective measurement of airflow limitation. It is a noninvasive and readily available test." (GOLD 2022)

The GOLD 2022 recommendations are an update of the GOLD 2015 guideline recommendation that was provided in the last submission:

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

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

Evidence to support the ACP, ACCP, ATS, ERS guideline recommendation is based on the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) workgroup and is classified as: **Strong-recommendation based on moderate-quality evidence.**

The following description of the definition of the grading of the evidence is taken from our prior submission, as it has not changed since:

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

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

This was not provided in the prior submission, so is added here:

Evidence is considered to be high quality when it is obtained from RCTs without important limitations or overwhelming evidence from observational studies. Evidence is low-quality when it is obtained from observational studies or case series. Strong recommendations can be made with high-, moderate-, or low-quality evidence. If strong recommendations are based on moderate or high quality evidence, then they can be applied to most patients in most circumstances without reservation. Strong recommendations based on low-quality evidence could change when higher quality evidence becomes available.

Weak recommendations can be made based on moderate- or high-quality evidence in cases where the benefits must be closely balanced with the risks and the burdens and the best action may differ depending on circumstances or patients' or societal values. Weak recommendations based on low-quality evidence obtained from observational studies or case series indicate that there is uncertainty in the estimates of benefits, risks and burden or they may be closely balanced; other alternatives may be equally reasonable.

When the balance of benefits and risks cannot be determined due to evidence that is conflicting, poor quality, or lacking, then it is decided that there is insufficient information to make a recommendation. (Qaseem et al 2010).

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

See 1a.05 above

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

See 1a.06 above

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

The following is copied from our prior submission, as it has not changed: "The quantity of studies reviewed in the ACP/ACCP/ATS/ERS guideline was not stated, but the guideline paper references 62 articles. This guideline is based on a targeted literature update from March 2007 to December 2009 to evaluate the evidence and update the 2007 ACP clinical practice guideline on diagnosis and management of stable COPD."

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

The following is copied from our prior submission with only minor edits for clarity, as the substance has not changed:

Targeted use of spirometry for diagnosis of airflow obstruction is beneficial for patients with respiratory symptoms, particularly dyspnea. In symptomatic patients, spirometry is helpful for determining whether the symptoms are due to respiratory disease or other conditions and to ensure appropriate therapy.

Existing evidence does not support the use of spirometry to screen for airflow obstruction in individuals without respiratory symptoms, including those with current or past exposure to risk factors for COPD. In asymptomatic individuals who have spirometric evidence of airflow obstruction, there is insufficient evidence to support treatment as it does not change annual rate of FEV1 decline or prevent of symptoms. No evidence from RCTs supports treating asymptomatic individuals, with or without risk factors for airflow obstruction, who do not have spirometric evidence of airflow obstruction. In addition, evidence does not show any independent benefit of obtaining and providing spirometry results on success rates in smoking cessation. No study evaluated the use of periodic spirometry after initiation of therapy to monitor ongoing disease status or modify therapy.

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

No specific harms of spirometry were identified in the guideline.

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

We have identified several studies published since our last submission that examine the use of spirometry for the diagnosis of COPD. These studies are summarized briefly here and added to relevant sections of the Evidence Section. None of these studies change the conclusions from the systematic review performed as part of the 2011 ATS Guidelines and the most recent GOLD statement, both of which endorse the use of spirometry to confirm airflow obstruction among patients with respiratory symptoms and risk factors for COPD.

Studies that demonstrate continued underuse of spirometry for the diagnosis of COPD:

One recent study assessed patients hospitalized within VA for COPD and found that only a little more than half (54.2%) had spirometry performed within a year before or after hospitalization. While the study did not report the proportion of patients that had ever had spirometry performed, it does provide continued evidence of overall underuse of the test, even among high risk patients (Rodwin et al., 2022)

Several other studies have focused on the discordance of spirometry findings and clinician diagnosis, resulting in both under- and over-diagnosis of COPD:

Diab and colleagues published a concise clinical review in 2018 that highlighted the prevalence of both under- and over-diagnosis of COPD (Diab N et al., 2018). The authors estimated that about 70% of COPD world-wide may be underdiagnosed. In contrast, 30-60% of patients are over-diagnosed, with no evidence of airflow obstruction on spirometry despite a clinician diagnosis of COPD.

Among 1403 participants in the Canadian Obstructive Lung Disease study, 14% had undiagnosed COPD, 5% were over-diagnosed, and only 4% had correctly diagnosed COPD (Gershon AS et al, 2018). Compared to patients without COPD, those with over-diagnosed COPD had higher healthcare utilization (hospitalization, ED and outpatient visits). Patients with undiagnosed COPD had higher rates of hospitalization than those without COPD.

Among 16,010 current and former smokers aged 55-77 participating in a lung cancer screening study, 1 in 5 patients were found to have undiagnosed COPD, defined as respiratory symptoms, no prior COPD diagnosis, and the presence of airflow obstruction on spirometry (Tisi et al., 2022)

Among 21,242 participants in the Canadian Longitudinal Study on Aging, researchers found significant discordance between self-reported COPD and spirometry findings, with 4% of participants representing under-diagnosis (no self-reported COPD, but airflow obstruction) and 4% having over-diagnosis (self-reported COPD, but no airflow obstruction). Participants with confirmed, under- and over-diagnosed COPD had higher risk of respiratory symptoms and health care utilization than participants with no self-reported COPD and no airflow obstruction (Farooqi MA et al, 2022).

Studies that address the potential of expanding the use of spirometry to screen asymptomatic patients with COPD:

Bhatt and colleagues examined lung function and clinical data from nine U.S general population cohorts to determine the burden of subclinical airflow obstruction (SAO) and develop a probability score of for SAO to in form detection programs (Bhatt et al., 2022). Of 33, 546 patients, 13.2% had SAO, which was associated with a 3-fold higher incidence of hospitalizations and death due to COPD. The probability score, based on demographic variables and smoking history, was well calibrated and showed excellent discrimination. While the score may prove useful in future targeted screening programs, this has not yet been tested.

In 2016, the US Preventative Services Task Force reviewed the evidence for and against screening asymptomatic individuals for COPD and determined there was no evidence that screening for COPD improved outcomes. In 2022, they published an updated recommendation statement based on an updated systematic review and reiterated their position against screening asymptomatic individuals.

As a result, our measure remains focused on the use of spirometry to confirm a diagnosis of COPD, rather than its use as a screening tool.

References:

Tisi S, Dickson JL, Horst C, et al. Detection of COPD in the SUMMIT Study Lung Cancer Screening Cohort using Symptoms and Spirometry. Eur Respir J. 2022 Jul 26; online ahead of print (https://doi.org/10.1183/13993003.00795-2022).

Rodwin, BA, DeRycke, EC, Han, L. et al. Characteristics Associated with Spirometry Guideline Adherence in VA Patients Hospitalized with Chronic Obstructive Pulmonary Disease. J Gen Intern Med. 14 Oct 2022.

Farooqi MAM, Ma J, Ali MU, et al. Prevalence and burden of COPD misclassification in the Canadian Longitudinal Study on Aging (CLSA). BMJ Open Resp Res 2022;9(1):e001156. doi:10.1136/bmjresp-2021-001156

Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet LP, and Aaron SD. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018 Nov 1;198(9):1130-1139.

Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, Tan W, To T for the Canadian Respiratory Research Network. Health services burden of undiagnosed and overdiagnosed COPD. Chest 2018; 153(6):1336-1346.

Bhatt SP, Balte PP, Schwartz JE, Jaeger BC, et al. Pooled cohort probability score for subclinical airflow obstruction. Ann Am Thorac Soc. 2022 Aug;19(8):1294-1304.

US Preventive Services Task Force, Siu AL, Bibbins-Domingo K, Grossman DC, et al. Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force recommendation statement. JAMA. 2016;315(13): 1372-1377

US Preventive Services Task Force, Mangione CM, Barry MJ, Nicholson WK, et al. Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force reaffirmation recommendation statement. JAMA. 2022 May 10;327(18):1806-1811.

[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]

Not applicable – evidence is from a systematic review performed as part of a clinical guideline.

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]

Targeted use of spirometry for diagnosis of airflow obstruction is beneficial for patients with respiratory symptoms, particularly dyspnea. In symptomatic patients, spirometry is helpful for determining whether the symptoms are due to COPD, other respiratory disease or non-pulmonary conditions and to ensure appropriate therapy. Confirming airflow obstruction with spirometry helps to identify who will benefit from treatment for COPD, including use of inhaled bronchodilators, long-term oxygen and/or pulmonary rehabilitation.

Existing evidence does not support the use of spirometry to screen for airflow obstruction in individuals without respiratory symptoms, including those with current or past exposure to risk factors for COPD. In asymptomatic individuals who have spirometric evidence of airflow obstruction, there is insufficient evidence to support treatment as it does not change annual rate of FEV1 decline or prevent of symptoms. No evidence from RCTs supports treating asymptomatic

individuals, with or without risk factors for airflow obstruction, who do not have spirometric evidence of airflow obstruction. In addition, evidence does not show any independent benefit of obtaining and providing spirometry results on success rates in smoking cessation. No study evaluated the use of periodic spirometry after initiation of therapy to monitor ongoing disease status or modify therapy.

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]

The following information includes information provided in the last submission:

"The ACP/ACCP/ATS/ERS guideline panel included representatives from each of the 4 collaborating organizations, and the resulting guideline represents an official and joint clinical practice guideline from those organizations. The guideline panel communicated via conference calls and e-mails. The members reached agreement and resolved any disagreements through facilitated discussion. The final recommendations were approved by unanimous vote. The key questions and scope for the guideline were developed with input from the joint guideline panel. Evidence reviews and tables were presented to the guideline panel for review and comments. The guideline panel evaluated the recommendations on the basis of the evidence."

Additional information available in the guideline methods is added for clarity with this submission:

"The Minnesota Evidence-based Practice Center performed an updated literature search that included studies from MEDLINE published between March 2007 and December 2009. Additional background material reviewed by the guideline panel included the 2007 systematic evidence review by Wilt and colleagues and the 2004 Agency for Healthcare Research and Quality–sponsored Minnesota Evidence-based Practice Center evidence report. The literature search focused on evidence for the value of spirometry for screening or diagnosis of COPD; the efficacy and comparative effectiveness of management strategies, such as inhaled monotherapies (anticholinergics, long-acting beta-agonists, or corticosteroids), combination therapies, and pulmonary rehabilitation programs, for patients with COPD. For diagnostic accuracy of the physical examination and spirometry, the guideline developers used an updated systematic review from 2008, because the guideline panel agreed that there is no reason to suspect that diagnostic accuracy of the physical examination or spirometry would have changed since the ACP guideline was published in 2007."

[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

Prior Submission:

Qaseem A, Snow V, Shekelle P, Sherif K, Wilt TJ, Weinberger S, et al; Clinical Efficacy Assessment Subcommittee of the American College of Physicians. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline from the American College of Physicians. Ann Intern Med. 2007;147:633-8.

Wilt TJ, Niewoehner D, MacDonald R, Kane RL. Management of stable chronic obstructive pulmonary disease: a systematic review for a clinical practice guideline. Ann Intern Med. 2007;147:639-53.

Agency for Healthcare Research and Quality. Use of Spirometry for Case Finding, Diagnosis, and Management of Chronic Obstructive Pulmonary Disease. Rockville, MD: Agency for Healthcare Research and Quality; 2005. Report no. 290-02-0009.

Amir Qaseem, MD, PhD, MHA; Timothy J. Wilt, MD, MPH; Steven E. Weinberger, MD; Nicola A. Hanania, MD, MS; Gerard Criner, MD; Thys van der Molen, PhD; Darcy D. Marciniuk, MD; Tom Denberg, MD, PhD; Holger Schu¨ nemann, MD, PhD, MSc; Wisia Wedzicha, PhD; Roderick MacDonald, MS; and Paul Shekelle, MD, PhD, for the American College of Physicians, the American College of Chest Physicians, the American Thoracic Society, and the European Respiratory Society* Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from

the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Ann Intern Med*. 2011;155:179-191.

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

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

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

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

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

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

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

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

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

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

Additional Citations from Current Submission:

Tisi S, Dickson JL, Horst C, et al. Detection of COPD in the SUMMIT Study Lung Cancer Screening Cohort using Symptoms and Spirometry. Eur Respir J. 2022 Jul 26; online ahead of print (https://doi.org/10.1183/13993003.00795-2022).

Rodwin, BA, DeRycke, EC, Han, L. et al. Characteristics Associated with Spirometry Guideline Adherence in VA Patients Hospitalized with Chronic Obstructive Pulmonary Disease. J Gen Intern Med.14 Oct 2022.

Farooqi MAM, Ma J, Ali MU, et al. Prevalence and burden of COPD misclassification in the Canadian Longitudinal Study on Aging (CLSA). BMJ Open Resp Res 2022;9(1):e001156. doi:10.1136/bmjresp-2021-001156

Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet LP, and Aaron SD. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018 Nov 1;198(9):1130-1139.

Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, Tan W, To T for the Canadian Respiratory Research Network. Health services burden of undiagnosed and overdiagnosed COPD. Chest 2018; 153(6):1336-1346.

Bhatt SP, Balte PP, Schwartz JE, Jaeger BC, et al. Pooled cohort probability score for subclinical airflow obstruction. Ann Am Thorac Soc. 2022 Aug;19(8):1294-1304.

US Preventive Services Task Force, Siu AL, Bibbins-Domingo K, Grossman DC, et al. Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force recommendation statement. JAMA. 2016;315(13): 1372-1377

US Preventive Services Task Force, Mangione CM, Barry MJ, Nicholson WK, et al. Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force reaffirmation recommendation statement. JAMA. 2022 May 10;327(18):1806-1811.

[Response Ends]

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

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

Current Submission: *The following rationale contains updated information from our recent literature search -:*

"Despite major efforts to broadly disseminate the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines and use of COPD performance measures across different specialty societies, COPD remains underdiagnosed and misdiagnosed (Collins et al., 2015; Perez et al., 2011; Diab et al., 2018; Gershon et al,, 2018; Tisi et.al 2022; Farooqi et al, 2022). Although spirometry use has increased, it remains underutilized to confirm airflow obstruction and accurately diagnose COPD (CDC, 2012; Nishi et al., 2013; Rodwin et al., 2022). Studies show proper COPD diagnosis with spirometry is done on just over half of patients in the US and Canada (Boulet et al., 2013; Bourbeau et al., 2008; Collins et al., 2015; Nishi et al., 2013; Perez et al., 2011; Yu et al., 2013) and ranges from 10-48% in the Asia-Pacific region, Africa, eastern Europe, and Latin America (Aisanov et al., 2012). World-wide, as many of 70% of patients with COPD may be underdiagnosed, while 30-60% of patients are over-diagnosed (Diab et. al, 2018) A study of physician-diagnosed COPD patients hospitalized for exacerbations found that 22% of patients did not have COPD upon spirometry testing (Prieto Centurion, et al., 2012). Treatment of COPD without accurate diagnosis and understanding of true etiology of symptoms results in patients not receiving medication that would improve symptoms and quality of life, prevent exacerbations and reduce costly use of emergency and hospital services. Patients may be exposed to adverse effects of unneeded medication and or delays in true diagnosis and management of another condition increasing overall cost of care (Boulet et al., 2013; Bourbeau et al., 2008; CDC, 2012; Collins et al., 2015; Joo et al., 2011). Several recent studies emphasize the association between both under- and over- diagnosis of COPD with increased respiratory symptoms and health care utilization (Gershon et al, 2018; Farooqi et al, 2022). We believe this measure will continue to increase appropriate spirometry use to assist physicians in the accurate diagnosis and treatment of patients with COPD, improving patient management and reducing total costs of COPD."

Additional New Citations:

1. Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet LP, and Aaron SD. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018 Nov 1;198(9):1130-1139.
2. Farooqi MAM, Ma J, Ali MU, et al. Prevalence and burden of COPD misclassification in the Canadian Longitudinal Study on Aging (CLSA). BMJ Open Resp Res 2022;9(1):e001156. doi:10.1136/ bmjresp-2021-001156
3. Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, Tan W, To T for the Canadian Respiratory Research Network. Health services burden of undiagnosed and overdiagnosed COPD. Chest 2018; 153(6):1336-1346.
4. Rodwin, BA, DeRycke, EC, Han, L. et al. Characteristics Associated with Spirometry Guideline Adherence in VA Patients Hospitalized with Chronic Obstructive Pulmonary Disease. J Gen Intern Med.14 Oct 2022.
5. Tisi S, Dickson JL, Horst C, et al. Detection of COPD in the SUMMIT Study Lung Cancer Screening Cohort using Symptoms and Spirometry. Eur Respir J. 2022 Jul 26; online ahead of print (https://doi.org/10.1183/13993003.00795-2022).

Prior Citations:

1. Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and

management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.

2. Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.
3. Bourbeau J, Sebaldt RJ, Day A, Bouchard J, Kaplan A, Hernandez P, Rouleau M, et al. Practice patterns in the management of chronic obstructive pulmonary disease in primary practice: the CAGE study. Can Respir J. 2008 JanFeb:15(1):13-9.
4. Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated healthcare resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.
5. Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.
6. Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.
7. Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.
8. Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.
9. Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.
10. Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

Previous 2015 Submission

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

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

Citations:

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

2. Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.
3. Bourbeau J, Sebaldt RJ, Day A, Bouchard J, Kaplan A, Hernandez P, Rouleau M, et al. Practice patterns in the management of chronic obstructive pulmonary disease in primary practice: the CAGE study. Can Respir J. 2008 Jan-Feb:15(1):13-9.
4. Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease and associated health-care resource use - North Carolina, 2007 and 2009. MMWR Morb Mortal Wkly Rep. 2012 Mar 2;61(8):143-6.
5. Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.
6. Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.
7. Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.
8. Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.
9. Prieto Centurion V, Huang F, Naureckas ET, Camargo CA Jr, Charbeneau J, Joo MJ, Press VG, Krishnan JA. Confirmatory spirometry for adults hospitalized with a diagnosis of asthma or chronic obstructive pulmonary disease exacerbation. BMC Pulm Med. 2012 Dec 7;12:73.
10. Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.

[Response Begins]

Information regarding performance of spirometry is publicly available for recent years for NQF measure 0577. Although there are important differences between 0577 and our measure (0091) (summarized in Section 5.05), the two are similar enough that a gap in this measure likely reflects a similar gap in our measure. This is further supported by the literature summarized in section 1b.03.

Spirometry testing rate by year among patients with COPD

Year	Commercial HMO	Commercial PPO	Medicaid HMO	Medicare PPO	Medicare
2020	37.4%	37.2%	26.8%	29.8%	31.5%
2019	41.5%	39.1%	30.4%	*	*
2018	41.7%	39.9%	31%	34%	34.2%
2017	41.6%	39.6%	31.6%	34.2%	33.5%
2016	42%	40.5%	31.6%	34.9%	35%

Year	Commercial HMO	Commercial PPO	Medicaid HMO	Medicare PPO	Medicare
2015	43.4%	41.1%	31%	36.3%	36.6%

* Cell intentionally left empty

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

https://www.ncqa.org/hedis/measures/use-of-spirometry-testing-in-the-assessment-and-diagnosis-of-copd/

Information from prior 2015 submission copied here:

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

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

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

Average performance rate:

- 2010 - 56.0%
- 2011 - 68.3%
- 2012 - 69.4%
- 2013 - 53.4%
- 2014 - 67.1%

Source: Timothy Jackson, CMS.

_____

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

CMS Physician Quality Reporting Initiative/System:

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

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

- 10th percentile: 4.17%
- 25th percentile: 17.39%
- 50th percentile: 51.45%
- 75th percentile: 83.33%
- 90th percentile: 94.85%
- Exception rate: 2.5%

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

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.

[Response Begins]

Additional information in current submission:

World-wide, as many of 70% of patients with COPD may be underdiagnosed, while 30-60% of patients are over-diagnosed (Diab et. All, 2018)

Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet LP, and Aaron SD. Underdiagnosis and Overdiagnosis of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018 Nov 1;198(9):1130-1139.

Prior response from 2015 submission copied here:

Studies show proper COPD diagnosis with spirometry is done on just over half of patients in the US and Canada (Boulet et al., 2013; Collins et al., 2015; Nishi et al., 2013; Perez et al., 2011; Yu et al., 2013,) and globally ranges from 6.5% in China to 59% in Sweeden with a mean of 26% in the Asia-Pacific region, Africa, eastern Europe, and Latin America (Aisanov et al., 2012; Yu et al., 2013).

Citations:

1. Aisanov Z, Bai C, Bauerle O, Colodenco FD, Feldman C, Hashimoto S, Jardim J, Lai CK, Laniado-Laborin R, Nadeau G, Sayiner A, Shim JJ, Tsai YH, Walters RD, Waterer G. Primary care physician perceptions on the diagnosis and management of chronic obstructive pulmonary disease in diverse regions of the world. Int J Chron Obstruct Pulmon Dis. 2012;7:271-82.
2. Boulet LP, Bourbeau J, Skomro R, Gupta S. Major care gaps in asthma, sleep and chronic obstructive pulmonary disease: a road map for knowledge translation. Can Respir J. 2013 Jul-Aug;20(4):265-9.
3. Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among Veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015 Feb;147(2):369-76.
4. Nishi SP, Wang Y, Kuo YF, Goodwin JS, Sharma G. Spirometry use among older adults with chronic obstructive pulmonary disease;1999-2008. Ann Am Thorac Soc. 2013 Dec:10(6):565-73.
5. Perez X, Wisnivesky JP, Lurslurchachai L, Kleinman LC, Kronish IM. Barriers to adherence to COPD guidelines among primary care providers. Respir Med. 2012 Mar;106(3):374-81.
6. Yu WC, Fu SN, Tai EL, Yeung YC, Kwong KC, Chang Y, Tam CM, Yiu YK. Spirometry is underused in the diagnosis and monitoring of patients with chronic obstructive pulmonary disease (COPD). Int J Chron Obst Pulmon Dis. 2013;8:389-95.

[Response Ends]

1b.04. 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.

[Response Begins]

Our response is unchanged from our prior submission in 2015:

"We are not aware of disparities data from this measure as specified. Please see 1b.05 for a summary of our findings in the literature regarding disparities."

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, 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 above.

[Response Begins]

Additional information in Current Submission:

***Disparities in COPD Diagnosis*:**

Mamary and colleagues examined the influence of race, gender, and severity of airflow obstruction on prevalence of COPD diagnosis at enrollment into COPDGene, a US multicenter cohort study designed to assess genetic susceptibility to COPD. Regardless of severity of airflow obstruction, African-Americans were less likely to have a prior COPD diagnosis at enrollment as compared to non-Hispanic whites. Women had a higher odds of prior COPD diagnosis than men regardless of severity of airflow obstruction. This study highlights gender and race disparities in the diagnosis of COPD. Since the diagnosis of COPD depends on spirometry, this study might be evidence of a disparity in performing spirometry on at-risk individuals.

Mamary AJ, Stewart JI, Kinney GL, et al. Race and gender disparities are evident in COPD underdiagnosis across all severities of measured airflow obstruction. Chronic Obstr Pulm Dis. 2018; 5(3):177-184.

***Additional studies of the association of gender and race with diagnosis of COPD*:**

A recent study examined difference in emphysema prevalence on CT scans among Black and White participants with normal spirometry participating in the Coronary Artery Risk Development in Young Adults (CARDIA) study (Liu et al., 2022). The study found a higher prevalence of emphysema among Black participants with race-specific FEV1 within the normal % predicted range. When race-neutral equations were used, these differences were attenuated among men and eliminated among women. The study concluded that reliance on spirometry alone to "differentiate between lung health and lung disease" may underestimate impaired respiratory health and exacerbate racial disparities. While the study makes an important contribution to the literature and generates discussion as to how and if CT scan findings should be incorporated into diagnostic criteria, the current guidelines remain unchanged and require spirometry to confirm COPD diagnosis.

Liu GY, Khan SS, Colangelo LA et al. "Comparing racial differences in emphysema prevalence among adults with normal spirometry: a secondary data analysis of the CARDIA lung study." Ann Intern Med. 2022 Aug;175(8):1118-1125.

Prior Response from 2015 Submission copied here:

"Studies have been done to show associations between education level and income and outcomes related to COPD (Eisner et al., 2011; Holt et al., 2011). Studies also show association between gender and race on the incidence/severity of COPD (Bruse et al., 2011; Diaz et al., 2014; Foreman et al., 2011; Han et al., 2011). However, few research studies have been conducted to show disparities in use of spirometry.

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

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

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

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

Citations:

1. Brehm JM, Celedón JC. Chronic obstructive pulmonary disease in Hispanics. Am J Respir Crit Care Med. 2008 Mar 1;177(5):473-8.
2. Bruse S, Sood A, Petersen H, Liu Y, Leng S, Celedón JC, Gilliland F, Celli B, Belinsky SA, Tesfaigzi Y. New Mexican Hispanic smokers have lower odds of chronic obstructive pulmonary disease and less decline in lung function than non-Hispanic whites. Am J Respir Crit Care Med. 2011 Dec 1;184(11):1254-60.
3. Celedón JC, Roman J, Schraufnagel DE, Thomas A, Samet J. Respiratory health equality in the United States. The American thoracic society perspective. Ann Am Thorac Soc. 2014 May;11(4):473-9.
4. Diaz AA, Come CE, Mannino DM, Pinto-Plata V, Divo MJ, Bigelow C, Celli B, Washko GR. Obstructive lung disease in Mexican Americans and non-Hispanic whites: an analysis of diagnosis and survival in the National Health and Nutritional Examination Survey III Follow-up Study. Chest. 2014 Feb;145(2):282-9.
5. Eisner MD, Blanc PD, Omachi TA, Yelin EH, Sidney S, Katz PP, Ackerson LM, Sanchez G, Tolstykh I, Iribarren C. Socioeconomic status, race and COPD health outcomes. J Epidemiol Community Health 2011;65:26–34.
6. Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, et al. Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene Study. Am J Respir Crit Care Med. 2011 Aug 15;184(4):414-20.
7. Ghattas C, Dai A, Gemmel DJ, Awad MH. Over diagnosis of chronic obstructive pulmonary disease in an underserved patient population. Int J Chron Obstruct Pulmon Dis. 2013;8:545-9.
8. Han MK, Curran-Everett D, Dransfield MT, Criner GJ, Zhang L, Murphy JR, Hansel NN, DeMeo DL, Hanania NA, Regan EA, Make BJ, Martinez FJ, Westney GE, Foreman MG; COPDGene Investigators. Racial differences in quality of life in patients with COPD. Chest. 2011 Nov;140(5):1169-76.
9. Holt JB, Zhang X, Presley-Cantrell L, Croft JB. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. Int J Chron Obstruct Pulmon Dis. 2011; 6 321–328.
10. Joo MJ, Au DH, Fitzgibbon ML, McKell J, Lee TA. Determinants of spirometry use and accuracy of COPD diagnosis in primary care. J Gen Intern Med. 2011 Nov;26(11):1272-7.

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]

No

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]

No changes

[Response Ends]

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

COPD: Spirometry Evaluation

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

Percentage of patients aged 18 years and older with a diagnosis of COPD who had spirometry results documented.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Respiratory

Respiratory: Chronic Obstructive Pulmonary Disease (COPD)

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

Diagnosis

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Adults (Age >= 18)

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Clinician: Group/Practice

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Outpatient Services

[Response Ends]

sp.09. 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. If no URL is available, indicate "none available".

[Response Begins]

The specifications for this measure are included within this form.

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

No data dictionary/code table – all information provided in the submission form

[Response Ends]

sp.13. State the numerator.

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.

[Response Begins]

Patients with documented spirometry results in the medical record (FEV1 and FEV1/FVC)

[Response Ends]

sp.14. Provide details needed to calculate the numerator.

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 sp.11.

[Response Begins]

Numerator Quality-Data Coding Options for Reporting Satisfactorily

Numerator Instructions: Look for documentation of spirometry evaluation results in the medical record at any time in the past; do not limit the search to the reporting period.

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

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

OR

Spirometry Results not Documented for Medical, Patient, or System Reasons

Append a modifier (1P, 2P or 3P) to CPT Category II code 3023F to report documented circumstances that appropriately exclude patients from the denominator.

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

OR

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

OR

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

OR

Spirometry Results not Documented, Reason not Otherwise Specified

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

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

[Response Ends]

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

All patients aged 18 years and older with a diagnosis of COPD

[Response Ends]

sp.16. Provide details needed to calculate the denominator.

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 sp.11.

[Response Begins]

All Patients aged >= 18 years on date of encounter

AND
Diagnosis for COPD

ICD-9-CM [for use before 9/30/2014]:
491.0, 491.1, 491.20, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 493.20, 493.21, 493.22, 496

ICD-10-CM [for use after 10/1/2014]:
J41.0, J41.1, J41.8, J42, J43.0, J43.1, J43.2, J43.8, J43.9, J44.0, J44.1, J44.9

(Please see listing below for ICD-9/ICD-10 code definitions)

AND

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

ICD-9/ICD-10 code definitions:

ICD-9-CM [for use before 9/30/2014]:
491.0 – Simple chronic bronchitis
491.1 – Mucopurulent chronic bronchitis
491.20 – Obstructive chronic bronchitis without exacerbation
491.21 – Obstructive chronic bronchitis with (acute) exacerbation
491.22 – Obstructive chronic bronchitis with acute bronchitis
491.8 – Other chronic bronchitis

491.9 – Unspecified chronic bronchitis
492.0 – Emphysematous bleb
492.8 – Other emphysema
493.20 – Chronic obstructive asthma, unspecified
493.21 – Chronic obstructive asthma with status asthmaticus
493.22 – Chronic obstructive asthma with (acute) exacerbation
496 – Chronic airway obstruction, not elsewhere classified

ICD-10-CM [for use after 10/1/2014]:
J41.0 – Simple chronic bronchitis
J41.1 – Mucopurulent chronic bronchitis
J41.8 – Mixed simple and mucopurulent chronic bronchitis
J42 – Unspecified chronic bronchitis
J43.0 – Unilateral pulmonary emphysema [MacLeod´s syndrome]
J43.1 – Panlobular emphysema
J43.2 – Centrilobular emphysema
J43.8 – Other emphysema
J43.9 – Emphysema, unspecified
J44.0 – Chronic obstructive pulmonary disease with acute lower respiratory infection
J44.1 – Chronic obstructive pulmonary disease with (acute) exacerbation
J44.9 – Chronic obstructive pulmonary disease, unspecified

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Documentation of medical reason(s) for not documenting and reviewing spirometry results
Documentation of patient reason(s) for not documenting and reviewing spirometry results
Documentation of system reason(s) for not documenting and reviewing spirometry results

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

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 sp.11.

[Response Begins]

ATS continues to use the PCPI exception methodology that uses three categories of exception reasons for which a patient may be removed from the denominator of an individual measure: medical, patient and system reasons.

Exceptions are used to remove patients from the denominator of a performance measure when a patient does not receive a therapy or service AND that therapy or service would not be appropriate due to specific reasons; otherwise, the patient would meet the denominator criteria. Exceptions are not absolute, and the application of exceptions is based on clinical judgment, individual patient characteristics, or patient preferences. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a

medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions include medical reason(s), patient reason(s) or system reason(s) for not documenting spirometry results. Although this methodology does not require the external reporting of more detailed exception data, the ATS recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The ATS also conducts systematic review and analysis of exceptions data to identify practice patterns and opportunities for quality improvement.

For Claims:

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

Append a modifier (1P, 2P or 3P) to CPT Category II code 3023F to report documented circumstances that appropriately exclude patients from the denominator.

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

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

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

The PCPI performed validity testing on a sample of 123 patient encounters from a single site (presented in their 2012 comprehensive review). There was only 1 exception (0.81%). Upon review, this was found not to be a valid exception, so percentages cannot be provided.

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

[Response Begins]

We encourage the results of this measure to be stratified by race, ethnicity, primary language, and administrative sex.

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.23. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

1. Start with Denominator
2. Check Patient Age:
    a. If the Age is greater than or equal to 18 years of age on Date of Service and equals No during the measurement period, do not include in Eligible Patient Population. Stop Processing.
    b. If the Age is greater than or equal to 18 years of age on Date of Service and equals Yes during the measurement period, proceed to check Patient Diagnosis.
3. Check Patient Diagnosis:
    a. If Diagnosis of COPD as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Diagnosis of COPD as Listed in the Denominator equals Yes, proceed to check Encounter Performed.
4. Check Encounter Performed:
    a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Patient Population. Stop Processing.
    b. If Encounter as Listed in the Denominator equals Yes, include in the Eligible population.
5. Denominator Population:
    a. Denominator population is all Eligible Patients in the denominator. Denominator is represented as Denominator in the Sample Calculation listed at the end of this document. Letter d equals 80 patients in the sample calculation.
6. Start Numerator
7. Check Spirometry Results Documented and Reviewed:
    a. If Spirometry Results Documented and Reviewed equals Yes, include in Reporting Met and Performance Met.
    b. Reporting Met and Performance Met letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter a equals 40 patients in Sample Calculation.
    c. If Spirometry Results Documented and Reviewed equals No, proceed to Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results.
8. Check Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results:
    a. If Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.

b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b1 equals 10 patients in the Sample Calculation.
c. If Documentation of Medical Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results.

9. Check Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results:
   a. If Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.
   b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b2 equals 0 patients in the Sample Calculation.
   c. If Documentation of Patient Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results.
10. Check Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results:
    a. If Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results equals Yes, include in Reporting Met and Performance Exclusion.
    b. Reporting Met and Performance Exclusion letter is represented in the Reporting Rate and Performance Rate in the Sample Calculation listed at the end of this document. Letter b3 equals 0 patients in the Sample Calculation.
    c. If Documentation of System Reason(s) for Not Documenting and Reviewing Spirometry Results equals No, proceed to Spirometry Results Not Documented and Reviewed, Reason Not Specified.
11. Check Spirometry Results Not Documented and Reviewed, Reason Not Specified:
    a. If Spirometry Results Not Documented and Reviewed, Reason Not Specified equals Yes, include in Reporting Met and Performance Not Met.
    b. Reporting Met and Performance Not Met letter is represented in the Reporting Met in the Sample Calculation listed at the end of document. Letter c equals 20 patients in the Sample Calculation.
    c. If Spirometry Results Not Documented and Reviewed, Reason Not Specified equals No, include in Reporting Not Met.
12. Check Reporting Not Met
    a. If Reporting Not Met equals No, Quality Data Code or equivalent not reported. 10 patients have been subtracted from the reporting numerator in sample calculation.

´Sample Calculation´ referenced above can be found in Appendix 1

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.

• The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.

• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.

• When possible, units of measurement and patients within units should be randomly selected.

[Response Begins]

Not applicable. The measure does not require sampling or a survey.

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

The measure is specified for claims data; it is not specified for other data sources.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

No

[Response Ends]

2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

No

[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]

No

[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]

No additional risk adjustment analysis included

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- 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.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the 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 the 2021 Measure Evaluation Criteria and Guidance.

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.

2a. Reliability testing 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 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;

AND

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

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
- 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 non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

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).

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.

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.

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

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

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.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Claims

[Response Ends]

2a.02. 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).

[Response Begins]

From 2015 submission

"The data source for reliability testing that was performed is the Centers for Medicare & Medicaid Services (CMS) Medicare administrative claims database.

The testing was conducted by Mathematica Policy Research as a component of the 2012 Quality and Resource Use Report (QRUQ), part of the CMS Physician Feedback Reporting Program.

Citation:

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

This measure was also tested in 2012 by the PCPI to support NQF re-endorsement for the 2012 comprehensive review which has been provided previously with the 2015 submission and is copied again here:

"EHR Measure Validity

The measure was calculated using data collected using two different methods of collection:

- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

The data source was electronic health records in an ambulatory care setting.

The data sample came from 1 site representing an academic medical center located in an urban area.

The sample consisted of 123 patient encounters.

Data collected from patients seen between 01/01/2010-12/31/2011.

Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012."

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

01-2012 - 12-2012

The most recent reliability testing is the Mathematica report that was published in January 2014 and used claims data from 2012. Please indicate if more updated testing is needed.

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Clinician: Group/Practice

[Response Ends]

2a.05. List the measured entities 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.

[Response Begins]

From the 2015 submission:

"Testing and analysis included 2,064 groups of physicians with at least 25 eligible professionals (EPs) (average of 120 EPs per group). Of these, there were 693 groups of physicians with at least 100 EPs (average of 322 EPs). This group represents 30% of medical group practices with 25 or more EPs nationwide. Groups were included if they reported at least 20 eligible cases for the measure. The groups were distributed across all states, the District of Columbia, Guam and Puerto Rico. "

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

From the 2015 submission:

"Testing and analysis included 11,593,241 Medicare beneficiaries identified on claims associated with the groups described in 1.5. Beneficiaries attributed to groups with more than 25 EPs averaged 2,974 (standard deviation = 5,105). Approximately half (52%) of the groups were attributed fewer than 1,000 beneficiaries. Beneficiaries attributed to groups with more than 100 EPs averaged 7,077 (standard deviation = 7,842)."

[Response Ends]

2a.07. 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.

[Response Begins]

From the 2015 submission:

The data were used for reliability testing only. Face validity testing was done with a survey. Other analyses were not done or not applicable.

[Response Ends]

2a.08. List 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.

[Response Begins]

From the 2015 submission:

"Patients in the testing and analysis were Medicare beneficiaries. No other sociodemographic variables were available for analysis."

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing 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.

[Response Begins]

As noted in 2a.02: Reliability testing was performed using Centers for Medicare & Medicaid Services (CMS) Medicare administrative claims database.

The testing was conducted by Mathematica Policy Research as a component of the 2012 Quality and Resource Use Report, part of the CMS Physician Feedback Reporting Program. The information that is provided is obtained from the following publication.

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

We have provided details from the publication about how the reliability testing was performed. Additional details are not available.

From the 2015 submission:

The method of reliability testing as used by Mathematica Policy Research is described as:

"For each of these measures, reliability was estimated as a ratio of variation on performance between groups and the total variation (variation between groups and variation from measurement error):

"Reliability = Variation between groups/(Variation between groups + Variation within group)

"If a score is deemed highly reliable, we would expect that a group's performance rates would be very similar if performance were calculated on the basis of a random sample of the practice's beneficiaries.

"Reliability scores are represented on a continuum from zero to one. Scores closer to zero indicate lower reliability and scores closer to one indicate higher reliability. Although there is no universally agreed-upon minimum reliability threshold, reliability scores in the 0.40–0.70 range are often considered moderate, and scores greater than 0.70 are considered high."

[see 2a.02 for citation]

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Copied from the 2015 submission:

"As noted above, scores above 0.70 are considered high.

The reliability for this measure among groups with 25 or more EPs was 0.73.

The reliability for this measure among groups with 100 or more EPs was 0.83."

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Copied from the 2015 submission:

"We believe this measure remains reliable based on high reliability test scores and relatively large test sample size. We also believe that the measure is reliable across relatively small groups and relatively large groups."

[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

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)

[Response Ends]

2b.02. 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.

[Response Begins]

Copied from the 2015 submission:

"Face validity of the measure score as an indicator of quality was systematically assessed using the following approach:

After the measure was fully specified, the ATS Clinical Practice Committee was asked to rate their agreement with the following statement:

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

The rating scale used was 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

The 12 members of the ATS COPD Clinical Practice Committee were selected to serve as an expert panel:

Kevin L. Kovitz,, MD

Robert DeMarco, MD

Scott Manaker, MD

Michael Donahoe, MD

Omar Hussain, MD

Katina Nicolacakis, MD

Tom Gildea, MD

Steve G. Peters, MD

Kashif Hussain, MD

Stephen Hoffman, MD

Alan Plummer, MD

Mike Nelson, MD"

Additional validity testing was performed by PCPI as part of 2012 comprehensive review - copied here and provided with the 2015 submission:

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

"EHR Measure Validity

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

Data analysis included:

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

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

Face Validity

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

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

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

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

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

Copied from the 2015 submission:

The results of the expert panel rating of the validity statement include:

- N = 12
- Mean rating = 4.6
- Panelists that agree or strongly agree that this measure can accurately distinguish good and poor quality = 91.7%

Frequency distribution of ratings

1 – Strongly disagree	0
2	0
3 – Neither Agree nor Disagree	1
4	3
5 – Strongly Agree	8

Table displaying the frequency of distribution ratings from strongly disagree to strongly agree

Additional validity testing was performed by PCPI as part of 2012 comprehensive review - copied here and provided with the 2015 submission:

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity

This measure demonstrates substantial agreement when comparing the automated EHR report to visual inspection.

Reliability: N, % Agreement, Kappa

Numerator: 123, 86.89%, 0.7281 (0.6086-0.8476 CI)

Denominator: 123, 100%, kappa non-calculable (non-calculable CI)*

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

Face Validity

The results of the expert panel rating of the validity statement were as follows: N = 7; Mean rating = 4.86 and 100% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings

1 - 0 (Strongly Disagree)

2 - 0

3 - 0 (Neither Agree nor Disagree)

4 - 1

5 - 6 (Strongly Agree)

[Response Ends]

2b.04. Provide 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?)

[Response Begins]

Copied from the 2015 submission as remains unchanged:

"We believe this measure remains valid based on the degree of agreement by a panel of testers."

[Response Ends]

2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

Copied from the 2015 submission as remains unchanged:

"Analysis of the differences in performance rates was conducted through benchmarks. According to Mathematica Policy Research, 'Prior-year benchmarks were also computed for the claims-based quality indicators, and none of the measures differed significantly at the 5 percent level from the prior year benchmark. A weighted average (based on eligible cases) of performance for groups with 25 or more EPs serves as the benchmark for all groups of this size, whereas a comparable weighted average among groups with at least 100 EPs forms the benchmark for larger groups (100 or more EPs).'"

[see 2a.02 for citation]

Additional data from the 2012 comprehensive review testing form submitted by the AMA-PCPI: Copied here from the 2015 submission.

"CMS Physician Quality Reporting Initiative/System:

98,074 cases were reported on for the 2008 program, the most recent year for which data is available.

The following information is for the 2009 program, the only year for which such data is available.

Clinical Condition and Measure: #51 Spirometry Evaluation

Eligible Professionals: 212,885

Professionals Reporting: 1,841

% Professionals Reporting: 0.86%

Professionals Reporting >=80% of eligible instances: 737

% Professionals Reporting >=80% of eligible instances: 40.03%

CMS Physician Quality Reporting Initiative/System:

The inter-quartile range (IQR) was calculated to determine the variability of performance on the measure."

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

[Response Begins]

Copied from the 2015 submission:

"The percent of groups different than the benchmark ($p<0.05$) for this measure among groups with 25 or more EPs was 45.6%.

The percent of groups different than the benchmark ($p<0.05$) for this measure among groups with 100 or more EPs was 47.1%."

Additional data from the 2012 comprehensive review testing form submitted by the AMA-PCPI: Copied here from the 2015 submission.

"CMS Physician Quality Reporting Initiative/System:

Scores on this measure: N = 98,074; Mean = 54.30%,

10th percentile: 4.17%

25th percentile: 17.39%

50th percentile: 51.45%

75th percentile: 83.33%

90th percentile: 94.85%

The inter-quartile range (IQR) provides a measure of the dispersion of performance. The IQR is 65.94 and indicates that 50% of physicians have performance on this measure ranging from 17.39% and 83.33% and 10% of physicians have performance rates less than or equal to 4.17%.(1)"

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

[Response Ends]

2b.07. Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

Copied from the 2015 submission:

The proportion of groups statistically different than the benchmark suggests that there is variation across group performance.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

Missing data analysis was not conducted as part of the reliability testing performed by Mathematica.

Copied from the 2015 submission:

Missing data analysis was not conducted on this measure in this study.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

Copied from the 2015 submission:

Not available

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, 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 was conducted, justify the selected approach for missing data.

[Response Begins]

Copied from the 2015 submission:

Not available

[Response Ends]

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 eCQMs). 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.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. 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. Indicate what statistical analysis was used.

[Response Begins]

From 2015 submission: Not applicable

Also provided in the 2015 Submission and copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

The measure was calculated using data collected using two different methods of collection:

- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

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

Data analysis included:

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

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

From 2015 submission: Not applicable

Also provided in the 2015 Submission and copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

"EHR Measure Validity

This measure demonstrates substantial agreement when comparing the automated EHR report to visual inspection.

Reliability: N, % Agreement, Kappa

Numerator: 123, 86.89%, 0.7281 (0.6086-0.8476 CI)

Denominator: 123, 100%, kappa non-calculable (non-calculable CI)*

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

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

This measure demonstrates substantial agreement when comparing the automated EHR report to visual inspection.

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

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?

[Response Begins]

No updated testing has been performed. Please indicate if it is needed.

From 2015 submission:

"Exclusion analysis was not conducted on this measure in this study".

Also provided in the 2015 Submission and copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

"EHR Measure Validity

The data sample came from 1 site representing an academic medical center located in an urban area.

The sample consisted of 123 patient encounters.

Data collected from patients seen between 01/01/2010-12/31/2011.

Visual inspection of the medical record was performed between 02/06/2012 and 02/10/2012.

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

[Response Ends]

2b.17. Provide 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.

[Response Begins]

From 2015 submission: Not available

Also provided in the 2015 Submission and copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

Copied from 2012 comprehensive review testing form submitted by the AMA-PCPI:

EHR Measure Validity

Exception rate: 0.81%

Validity of exceptions was 0% agreement with a kappa of 0.0000*

*Due to the small sample size and the single exception found during manual abstraction, the resulting agreement rate and kappa statistic are low.

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

[Response Begins]

Although the number of exceptions was low in the abstracted records, they remain necessary as there may be medical reasons that spirometry cannot be performed, patients may choose to not undergo the procedure, or the testing may not be available in a healthcare system (e.g. spirometry was suspended in many health systems due to the COVID pandemic).

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

No risk adjustment or stratification

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

Not applicable.

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

[Response Ends]

2b.26. 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

Not applicable

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

[Response Ends]

Criterion 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.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

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-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in electronic claims

[Response Ends]

3.03. 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 data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A

[Response Ends]

3.06. 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.

[Response Begins]

We are not aware of any difficulties regarding data collection. As the measure is specified from claims data, there should be no problems with obtaining data, availability of data, patient confidentiality, or other fe asibility or implementation issues. All claims data has some degree of misclassification of diagnosis or missingness; this should not affect this particular measure any more than other claims based measures.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]

The measure is free to use.

[Response Ends]

Criterion 4: Use and Usability

4a. 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.

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

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 demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

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

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

The measure was previously part of Medicare's Physician Quality Reporting System (PQRS), which has now changed to the Merit Based Incentive Program (MIPS), part of CMS's Quality Payment Program. The measure is currently not part of the MIPS Quality Measures, though spirometry is required in order to meet another measure, use of long-acting inhaled bronchodilators in patients with COPD (NQF 0102). CMS is seeking input on their proposed MIPS Value Pathways for 2023. We are coordinating a response with several other professional societies and are scheduling a meeting with them to provide feedback. During this meeting, we plan to suggest that the spirometry measure is added back to the MIPS/MVP quality measures, given the importance of confirming accurate COPD diagnosis and continued widespread underuse of spirometry.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Quality Improvement (internal to the specific organization)

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

As noted above, the measure was previously part of Medicare's PQRS Program (which ended in 2016). The measure was initially part of the MIPS program as well, through Performance Year 2019. Starting in Performance Year 2020, the individual measure was dropped from the MIPS program because documentation of spirometry is a required component of another measure, the appropriate use of long-acting inhaled bronchodilators (NQF 102). We still believe that the measure has value as a stand-alone measure, given the widespread underuse of spirometry to confirm a diagnosis of COPD. If spirometry is not performed, patients are not eligible for the denominator for NF102, so the use of NQF102 neither ensures that spirometry is performed, nor assesses the frequency of its performance.

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

[Response Begins]

CMS has reached out to the American Thoracic Society, along with the Academy of Allergy, Asthma, and Immunology (AAAAI), American College of Chest Physicians (ACCP), American Lung Association (ALA) to provide multi-society input on the development of MIPS Value Pathways, which are intended to allow clinicians to report on quality measures in the fields of COPD, asthma, Sleep and general pulmonary. We are in the process of scheduling a meeting with them to provide overall feedback, which will include the importance of the inclusion of the spirometry measure in their MVP program. We are optimistic that the data support the inclusion of the spirometry measure in the program to promote improved diagnosis and care quality for COPD.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

[Response Begins]

CMS publicly reports Quality Payment Program (QPP) performance information for doctors, clinicians, groups and accountable care organizations (ACOs) on Medicare Care Compare Doctors and Clinicians profile pages and in the Provider Data Catalog (PDC). (Previously this information was reported on Physician Compare Profile pages and in the Physician Compare Downloadable Database).

CMS reports MIPS eligible clinicians' final scores and performance under each MIPS performance category, including for the spirometry measure through Performance Year 2019. While this information was previously reported one year at a time, CMS recently expanded the archive for Doctors and Clinicians in the PDC on Care Compare, allowing users to access historic MIPS program performance data dating back to the programs' inception in 2017 (including the spirometry measure).

More information available at:

https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/care-compare-dac-initiative

https://data.cms.gov/provider-data/topics/doctors-clinicians

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

See 4a.05

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

No specific feedback has been obtained, but we have had no problems reported by CMS.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

No specific feedback has been obtained, but we have had no problems reported by CMS.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

No specific feedback has been obtained, but we have had no problems reported by CMS.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

No specific feedback has been obtained, but we have had no problems reported by CMS.

[Response Ends]

4b. Usability

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

[Response Begins]

As highlighted in the Importance to Measure and Report section, there remains significant underuse of spirometry to confirm the diagnosis of COPD. For this reason, we are providing feedback to CMS about the importance of adding the measure back to its QPP MIPs value pathways.

The performance results from this measure can be easily used by clinicians to improve care, as there are minimal barriers to increasing the use of spirometry that would be faced by clinicians who seek to do so. Spirometry is reimbursed and is readily available in most practice settings and creates minimal burden to patients. Increased use of spirometry will improve care by minimizing the risk of misdiagnosis of COPD in patients who have other conditions that cause respiratory symptoms and by verifying the presence of airway obstruction in patients who do have COPD. In fact, one intervention shown to improve outcomes in COPD is only reimbursed in patients who meet spirometry criteria for reimbursement.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

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

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

We are not aware of any unexpected benefits related to this measure.

[Response Ends]

Criterion 5: Related and Competing Measures

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

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

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

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

N/A

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

No

[Response Ends]

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

[Response Begins]

These measures have distinct differences in their denominators and numerators. First, our measure is broader in denominator population, being for all patients age 18 years and older with a diagnosis of COPD, while 0577 is for patients age 40 years and older with a new diagnosis of COPD. Our measure is more consistent with COPD guidelines, which do not state an age to start using a spirometry evaluation; rather, spirometry should be used to assess all adults with COPD, not just adults with a new diagnosis of COPD. Second, our measure´s numerator is more flexible than 0577, allowing a spirometry evaluation anytime during the measurement period, rather than 0577´s requirement that spirometry be performed within 6 months of a new diagnosis of COPD. Our measure numerator is also specific to spirometry results, requiring both the FEV1/FVC values.

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

We believe our measure to be superior to the competing measure for reasons specified above: 1) our measure has a broader denominator population that is more consistent with COPD guidelines rather than just patients "newly" diagnosed;- 2) we allow spirometry to be performed any time in the past for confirmation of a diagnosis, rather than the requirement of it being performed only within a specified time period (please note, the time period listed in 5.05 inadvertently says within 6 months, but should have said within 2 years prior to the index episode date or 6 months after the episode start date). This is more consistent with standard clinical practice as many patients with COPD do not need regularly repeated spirometry, so even one performed several years prior is adequate to confirm COPD in the right clinical circumstances; and 3) our measure numerator requires both FEV1 and FVC values, making our validation of COPD based on the spirometry more accurate.

[Response Ends]