NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 1893

Corresponding Measures:

De.2. Measure Title: Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services

De.3. Brief Description of Measure: The measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR), defined as death from any cause within 30 days after the index admission date, for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD. CMS annually reports the measure for patients who are 65 years or older and enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals or are patients hospitalized in Veterans Health Administration (VA) facilities.

1b.1. Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for COPD. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Additionally, COPD mortality is a priority area for outcomes measure development, as it is a costly and common condition. Hospital mortality is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices. Furthermore, the measure will increase transparency for consumers and potentially lower the healthcare costs associated with mortality.

S.4. Numerator Statement: The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days from the date of admission for patients hospitalized with either a principal diagnosis of COPD or a principal diagnosis of respiratory failure with a secondary diagnosis of acute exacerbation of COPD.

S.6. Denominator Statement: This claims-based measure is used for a cohort of patients aged 65 years or older.

The cohort includes admissions for patients aged 65 years and older discharged from the hospital with a principal discharge diagnosis of COPD and with a complete claims history for the 12 months prior to admission. The measure is publicly reported by CMS for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details.

S.8. Denominator Exclusions: The mortality measures exclude index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
2. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
3. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort for each year.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data, Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jan 08, 2013 **Most Recent Endorsement Date:** Aug 03, 2016

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is paired with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization.

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

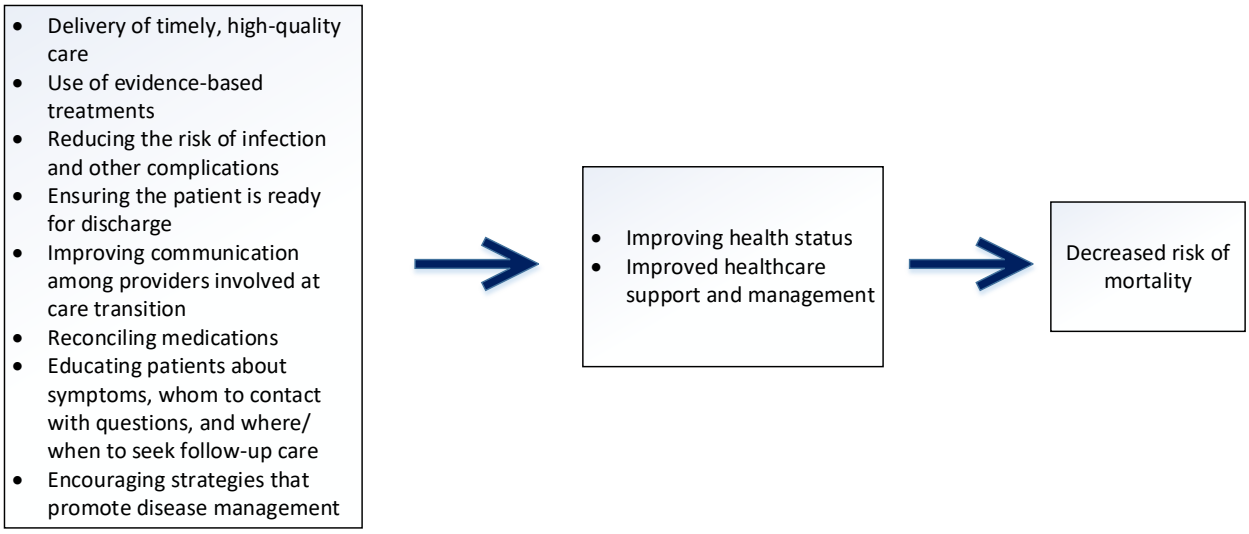
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- The developer provided a logic model that demonstrated how specific interventions have been associated with reduce COPD mortality.

Figure 1. COPD Mortality Logic Model



Changes to evidence from last review

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

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

Updates:

- Additional literature was provided by the developer that supported COPD as an important, common, high-cost and complex condition.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *Does the additional literature support the continued importance of COPD mortality is as an important outcome in older adults.*

Guidance from the Evidence Algorithm

Box 1 – Health outcome? (Yes) -> Box 2 – Is there one or more healthcare actions that can be taken to improve this measure? -> PASS

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Using data from July 1, 2016 to June 30, 2019 with Medicare claims and VA administrative data (n= 716,323 admissions from 4,642 hospitals), the three-year hospital-level risk-standardized mortality rates (RSMRs) had a mean of 8.4% and range from 5.1-13.6% in the study cohort.
- The median risk-standardized rate was 8.3%.

Disparities

- Among dual eligible (with Medicare and Medicaid insurance) in the same data, comparing the social risk proportion and the AHRQ SES index, there was a similar distribution of RSMR scores.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Is the disparities data provided sufficient?

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

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

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

- High rating, no concerns.
- Again, communication from the attending physician to the PCP remains a key process that is often done poorly
- Evidence is strong to measure differences in quality, and literature reviews support ongoing use of this measure
- Evidence to support
- I know of now applicable new studies, nor do I have suggestions on how the measure might be improved.
- Providers can perform at least one intervention to impact this measure and the additional evidence supports a continued focus on copd mortality in older adults
- The rating is Pass. Agree evidence is present that one or more interventions can be used to reduce COPD mortality.

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

- No concerns.
- Continued gap in performance
- 26% difference between high performing and low performing hospitals is meaningful but small Ns in each group warrant a little concern. Subgroup analysis shows minimal dual eligible differences or by neighborhood SES. However, bottom quartile may not reflect contemporary theories on how

neighborhoods impact health--might need more targeted definition of "severe" poverty at bottom 10-15%.

- High performance gap
- There appears to be an adequate performance gap, indicating the need to continue the measurement, but the data are 5 years old (2016). Data on disparities were quite limited.
- There are performance gaps in care and disparities data was sufficient
- From 2016 to 2019 range of COPD mortality range from 5.6% to 13.6% with 10th percentile around 7% and 90th percentile close to 10%. Agree that there is still room for improvement. Rating is High.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

SMP Rating:

- **R:** H-0; M-6; L-1; I-0 (Pass)
- **V:** H-2; M-5; L-0; I-0 (Pass)

Evaluators: NQF Scientific Methods Panel

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- Two types of reliability testing at the level of the performance measure score: 1) the intra-class correlation coefficient (ICC) using a split sample (i.e. test-retest) method, and 2) the facility-level reliability (signal-to-noise reliability).
- Split-Sample Reliability Results:
    - In 716,232 admissions over 3 years of data, this was split into two samples. ICCs were calculated for hospitals with 25 admission or more. Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of RSMR for each hospital was 0.477.
- Signal-to-Noise Results:
    - The median reliability was 0.72 with a range of 0.32 to 0.97 with the IQR of 0.54 (25^{th}) to 0.83 (75^{th}).

Validity

- The developer conducted empirical validity testing as well as a systematic assessment of face validity.
- Empirical validity results:
    - There was validation of the performance of the claims-based model and a medical records-based model with similar areas under the receiver operating characteristic (ROC) curve of 0.69 and 0.77, respectively, for the two models. The developer also estimated hospital-level RSMRs using hierarchical logistic regression administrative and medical record models then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital. The correlation coefficient of the standardized rates from the administrative and medical record models was 0.91 showing a strong correlation in rates calculated from the clinical and administrative models.
    - Two measures were the basis of comparison, the Hospital Star Rating Mortality group and the overall Hospital Star rating.
        - The correlation between COPD RSMRs and Star-Rating mortality score was -0.618, suggesting that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating mortality scores.
        - The correlation between COPD RSMRs and Star-Rating summary score was -0.165, suggesting that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating summary scores.
- Face validity results:
    - An 11-member TEP assessed the face validity of the measure. This was from the original submission. Of the TEP members who responded, 90% agreed (60% moderately or strongly agreed) that the measure will provide an accurate reflection of quality.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

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

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

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

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

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

- High reliability
- No concerns
- Reliability testing was strong. Estimates of signal-noise reliability were high
- Agree with moderate prelim rating
- none
- No comment, no discussion or vote needed
- I do not have concerns
- 2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?
- No
- No concerns
- No
- No
- no
- No comment. No discussion or vote needed.
- Agree that reliability is not as high as desired but is still acceptable.

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

- No
- No concerns
- No. Good correlations with related measures suggests this one is assessing a robust construct with face validity and construct validity
- No
- I wonder if some of those discharged with a diagnosis of COPD die in ways that are not reflected in the patient's EHR of claims data. For example, if someone dies at home within 30 days of admission, how is their death captured in the measure calculation. What about those who die from causes unrelated to COPD?
- No comment, no discussion or vote needed
- No concerns.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from

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

- N/A
- No concerns
- The overall risk adjustment approach, using statistical and clinical input, seems appropriate. However, the model may under adjust for functional impairment by not taking into account where patients were admitted from (e.g. admissions from home health care would indicate homebound and higher risk), admissions from SNF or long-term care may similarly indicate risk AND be disproportionately distributed among hospitals, especially those in rural areas. I also had a small concern with the AHRQ SES application; a substantial body of literature is emerging to show poorer outcomes for adults in low neighborhood SES settings. Many of those approaches re-calculate the count of hospitals that may be impacted (through better star ratings or fewer penalties) by adjusting for SES status. The overall average model metrics may not be affected, but there may be subsets of highly vulnerable hospitals that are missed with this approach. Lastly, neighborhood poverty may have a threshold where the impacts are most likely to be felt
- Agree with moderate prelim rating
- As noted above, some patients may die outside the ability of the EHR or claims data to capture that death - I think.
- No concerns
- I agree that the exclusions and the rationale for risk adjustment are both sound.

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

- No concerns.
- No concerns
- No
- Agree with moderate prelim rating
- As noted above, there may be death data that is not captured.
- No
- No concerns.

Criterion 3. Feasibility

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

3. **Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
    - All data elements are in defined fields in electronic claims

Questions for the Committee:

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

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- High feasibility. No concerns.
- No concerns
- Very feasible
- High feasibility
- none
- Measure is feasible
- No concerns. Rating High.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

Public Reporting: Hospital Compare https://www.medicare.gov/hospitalcompare/search.html?

Payment Program: Hospital Value Based Purchasing Program (HVBP)
https://www.qualitynet.org/inpatient/hvbp

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the

measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

Each hospital receives their measure results in the Spring of each calendar year through CMS's QualityNet website. The results are then publicly reported on CMS's Hospital Compare website in July of each calendar year. Since the measure is risk standardized using data from all hospitals, hospitals cannot independently calculate their score. Detailed reports are also provided each hospital, as well as a user guide and other resources.

Additional Feedback: The Yale CORE allows for feedback and questions directly from the hospitals to address specific questions about the measure specifications. Since the last endorsement cycle, we have reviewed more than 350 articles related to mortality following COPD admissions.

Some studies have argued that since HRRP implementation, mortality for some conditions (including COPD) has increased, suggesting a potential unintended consequence that readmission measures may be incentivizing hospitals to not readily admit patients with COPD, and as a result, mortality rates increased (Samarghandi et al., 2019). However, empiric findings and other studies have found no apparent increase in COPD mortality (Ni et al., 2016; MedPAC, 2018; Stensland., 2019).

Given the importance of this potential issue on patient outcomes, CMS commissioned an independent group to investigate whether there have been increases in mortality rates after HRRP implementation. CMS found through this investigation that no sufficient evidence exists to suggest that mortality has increased because of the HRRP readmission measures. CMS is committed to continuing to monitor trends in same-condition readmission and mortality rates through annual measure reevaluation and surveillance tasks.

***Questions for the Committee**:*

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

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

RATIONALE:

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

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

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

Improvement results The median hospital 30-day, all-cause, RSMR for the COPD mortality measure for the 3-year period between July 1, 2016 and June 30, 2019 was 8.3%. The median RSMR decreased by 0.7 absolute percentage points from July 2016-June 2017 (median RSMR: 8.6%) to July 2018-June 2019 (median: RSRR: 7.9%).

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

Unexpected findings (positive or negative) during implementation: The developer has reported no unintended consequences.

Potential harms: The developer has reported no potential harms.

Additional Feedback: The developer reports being committed to continuously monitoring this measure.

Questions for the Committee:

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

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

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

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

- Pass
- Needed to drive behaviors at time of transition
- Public reporting of data is ongoing, and currently being used in an accountability program.
- No concerns
- ok
- Measure is actionable and useful
- Currently being used and publicly reported. Rating Pass.

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

- No concerns
- No concerns
- Benefits outweigh harms, and no unintended consequences seem to have occurred with implementation
- Benefits > harms
- There seems to be concerns about hospitals not wishing to readmit patients, although I do not understand how this may impact negatively. One important factor seems to me to be the level of post-discharge care that keeps the patient alive for 30 days after the index admission. The hospital may not be well integrated with community support systems.
- Measure is relevant and useful
- Median mortality improved by 0.7% point so far. No identified unintended consequences. Rating high.

Criterion 5: Related and Competing Measures

Related or competing measures

0275 : Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)
0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization
0506 : Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization
1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
2888 : Accountable Care Organization Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions
3502 : Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure
3504 : Claims-Only Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure

Harmonization

Developer states that the measure is harmonized with existing measures.

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

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

- No concerns
- Measure is harmonized
- No concerns
- Several, developer states harmonized
- There are many related measures that are supposedly 'harmonized' with this one.
- No comment
- I find the explanations provided to differentiate this from other related measures to be satisfactory.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: **01/15/2021**

- Comment by: **American Medical Association**

The American Medical Association (AMA) appreciates the opportunity to comment on #468, Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization. We are disappointed to see the minimum measure score reliability results of 0.32 using a minimum case number of 25 patients and the intraclass correlation coefficients (ICC) was 0.477. We believe that measures must meet **minimum** acceptable thresholds of 0.7 for reliability and require higher case minimums to allow the overwhelming majority of hospitals to achieve an ICC of 0.6 or higher.

In addition, the AMA is extremely concerned to see that the measure developer used the recommendation to not include social risk factors in the risk adjustment models for measures that are publicly reported as outlined in the recent report to Congress by Assistant Secretary for Planning and Evaluation (ASPE) on Social Risk Factors and Performance in Medicare's Value-based Purchasing program (ASPE, 2020). We believe that while the current testing may not have produced results that would indicate incorporation of the two social risk

factors included in testing, this measure is currently used both for public reporting and value-based purchasing. A primary limitation of the ASPE report was that none of the recommendations adequately addressed whether it was or was not appropriate to adjust for social risk factors in the same measure used for more than one accountability purpose, which is the case for here. This discrepancy along with the fact that the additional analysis using the American Community Survey is not yet released must be addressed prior to any measure developer relying on the recommendations within this report.

We request that the Standing Committee evaluate whether the measure meets the scientific acceptability criteria.

Reference:

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

- Comment by: **Federation of American Hospitals**

The Federation of American Hospitals (FAH) appreciates the opportunity to comment on Measure #468, Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization. The FAH is concerned that even though the median reliability score was 0.72 for hospitals with at least 25 cases, reliability ranged from 0.32 to 0.97 and that the intraclass correlation coefficients (ICC) was 0.477. The FAH believes that the developer must increase the minimum sample size to a higher number to produce a minimum reliability threshold of sufficient magnitude (e.g. 0.7 or higher) and an ICC of 0.6 or higher.

In addition, the FAH is very concerned to see that the measure developer's rationale to not include social risk factors in the risk adjustment model was in part based on the recommendations from the report to Congress by Assistant Secretary for Planning and Evaluation (ASPE) on Social Risk Factors and Performance in Medicare's Value-based Purchasing program released in March of last year (ASPE, 2020). A fundament flaw within the ASPE report was the lack of any recommendation addressing how a single measure with multiple accountability uses should address inclusion of social risk factors as is the case with this measure, which is both publicly reported and included in the Hospital Value-Based Purchasing program. Regardless of whether the testing of social risk factors produced results that were sufficiently significant, the FAH believes that no developer should rely on the recommendations of this report until the question of how to handle multiple uses is addressed along with the additional analysis using the American Community Survey.

As a result, the FAH requests that the Standing Committee carefully consider whether the measure as specified meets the scientific acceptability criteria.

Reference:

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

- **Of the 1 NQF member who have submitted a support/non-support choice:**
  - **0 support the measure**
  - **1 does not support the measure**

Measure Number: 1893

Measure Title: Hospital, 30-day, all-cause, risk-standardized mortality rate (RSMR Following) following chronic obstructive pulmonary disease (COPD) hospitalization

Type of measure:

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

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

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☒ **Other**: Medicare Enrollment Data, VHA Administrative Data

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member #1: The 30 days begins with the index admission e.g., includes deaths in the hospital.

   Panel Member #4: No major concerns. However, I'm not completely clear what a "complete claims history for the 12 months prior to admission" is or how one can be sure the patient has it. If it is the same as "Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the index admission and Part A during the index admission, or those who are VA beneficiaries" then that should be indicated.

   Panel Member #5: The following exclusions are stated in the MIF, but not defined:

   -Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission

   -Discharged against medical advice (AMA). [p7]

   Panel Member #6: The specifications are precise.

Panel Member #7: None. RELIABILITY: TESTING

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

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: Appropriate methods used.

   Panel Member #3: The developers use 2 methods to estimate reliability: Method #1 involved randomly splitting the data into 2 samples in order to calculate 2 estimates per provider. The developers used two-way ANOVA to estimate ICC(2,1) as defined by Shrout & Fleiss (1979), and then converted this into an estimate of reliability for the full sample size (not half of the sample size). Method #2 involved estimating signal variance and provider-specific probabilities in a hierarchical model and then plugging these estimates into a formula for provider-specific reliability. The error variance is estimated using the variance of the logistic distribution, π^2/3. Method #1 seems relatively straightforward, and I have no questions. Method #2 is also reasonable, but I am unclear about the validity of the expression π^2/3 for estimating the error variance. This appears to be based on a model that views binary outcomes as the observed manifestation of an underlying latent continuous variable that follows a logistic distribution. If so, it seems that this formula could over-estimate reliability because it appears to estimate what reliability *would be* if we observed the underlying latent continuous variable. A dichotomized version of this variable contains less statistical information and could therefore yield less reliability.

   Panel Member #4: The developers used the following: ICC using a split sample (test-retest); and signal to noise (facility level reliability). The methods seem appropriate.

   Panel Member #5: The tests employed for reliability for this measure are appropriate.

   '…we estimated the overall measure score reliability by calculating the intra-class correlation coefficient (ICC) using a split sample (i.e. test-retest) method. Second, we estimated the facility-level reliability (signal-to-noise reliability).' [p9]

   Panel Member #6: The measure steward has added an analysis of facility-level reliability. The analysis is limited to hospitals with at least 25 COPD admissions. The familiar formula of Adams et al is used for reliability estimation.

   Panel Member #7: Both split-half reliability and signal-to-noise reliability are within current NQF guidance for appropriate reliability methods. The developer used hierarchical linear modeling to estimate facility level error variance.

   Panel Member #9: Methods were generally appropriate**.**
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3 Results indicate moderate reliability.

   Panel Member #3: The estimated reliability using split-sample testing (method #1) was 0.477. Provider-specific estimates using method #2 ranged from 0.32 to 0.97 (mean=0.65, median=0.72). As noted above, I wonder if method #2 might tend to over-estimate reliability.

   Panel Member #4: The developer stated the results were as follows:

   ICC using a split sample (test-retest)

   "the agreement between the two independent assessments of the RSMR for each hospital was 0.477"

The developers state " The split-sample reliability score of 0.477, discussed in the previous section, represents the lower bound of estimate of the true measure reliability."

My interpretation is that is lower than desired.

signal to noise (facility level reliability)

"The median reliability score was 0.72, ranging from 0.32 to 0.97. The 25th and 75th percentiles were 0.54 and 0.83, respectively."

The developers state: "Using the approach used by Adams et. al. and Yu et al., we obtained the median signal-to-noise reliability score of 0.72, which demonstrates substantial agreement."

My interpretation is that is this acceptable.

Panel Member #5: The test results for reliability were moderate for each test.

'Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the RSMR for each hospital was 0.477

We calculated the signal-to-noise reliability score for each hospital with at least 25 admissions* (see Table 2 below). The median reliability score was 0.72, ranging from 0.32 to 0.97. The 25th and 75th percentiles were 0.54 and 0.83, respectively' [p11]

Panel Member #6: Median reliability at the level of hospital was 0.72, with a range from 0.32 to 0.97. Table 2 suggests that approximately 20% of hospitals had reliability estimates that were less than 0.5.

Panel Member #7: The correspondence between the split-samples was 0.477, indicating that the reliable within hospital RSMR variance was 23%. Although the mean score for the signal-to-noise results was .72, reliability results for those scoring in the lowest quartile were .54 or lower. Sample sizes for the percentile categories reported would have been helpful to determine whether higher volume hospitals (as would be expected) had higher reliability estimates.

Panel Member #9: Results using the split-half method were marginal; results using the signal-to-noise method were better – at a threshold level for acceptable reliability.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   Panel Member #3: But see comments above.

   ☒ **No**

   Panel Member #7: Not reported

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

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

Panel Member #1: Testing results.

Panel Member #3: My moderate rating is based on the estimate of 0.477 from split sample testing. Also, results in 2b4 also shed light on reliability and suggest that the (shrunken) estimates are not dominated by sampling variation.

Panel Member #4: While the split half results are at the lower bond of acceptable, the ICC is at a moderate and acceptable range. I think, on balance, that there is sufficient reliability evidence.

Panel Member #5: Per Q2: 2 exclusions were not defined, which results in a "low rating" as noted immediately above: 'LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate, otherwise, reliability testing results were moderate (per Q7).

Panel Member #6: Moderate agreement in a refreshed split-sample analysis and median facility-level reliability roughly equal to 0.7 together suggest that the measure has limited value in hospitals with relatively low numbers of COPD admissions.

Panel Member #7: Although the methods and results are within current NQF guidance for reliability, the developer did not provide evidence suggesting that meaningful between facility differences could be established given error variance.

Panel Member #9: The measure score reliability for this measure was not impressive, but acceptable by recent standards for other endorsed measures.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #3: No concerns.

Panel Member #4: While it was not mentioned in the exclusion section, it was mentioned in **S.9. Denominator Exclusion Details**

that "if the patient has a sex other than 'male' or 'female'" an exclusion for the denominator? Is it because there might be too few people in this other category? I worry that this exclusion might leave out a vulnerable population.

Panel Member #5: No concerns

Panel Member #6: I have no concerns. The exclusions have a very small impact on available sample size.
Panel Member #7: While it was not mentioned in the exclusion section, it was mentioned in **S.9. Denominator Exclusion Details**

that "if the patient has a sex other than 'male' or 'female'" an exclusion for the denominator? Is it because there might be too few people in this other category? I worry that this exclusion might leave out a vulnerable population.

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

Submission document: Testing attachment, section 2b4.

Panel Member #1: No Concerns

Panel Member #3: No concerns.

Panel Member #4: The developers characterized the degree of variability by: Reporting the distribution of RSMRs; and providing the median odds ratio (MOR).

They state: The median odds ratio suggests a meaningful increase in the risk of mortality if a patient is admitted with COPD at a higher risk hospital compared to a lower risk hospital. A value of 1.26 indicates that a patient's risk of mortality is 26% greater in a higher risk hospital than a lower risk hospital. The variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for COPD. This evidence supports continued measurement to reduce the variation."

I have no concerns with this.

Panel Member #5: Of hospitals with an sufficient n to report"

3.3% were "worse" at the 95%CI

2.4% were "better" at the 95% CI [p34]

It's somewhat concerning that less than 6% of hospitals are identified as outliers through this measure.

Panel Member #6: I have no concerns. A median odds ratio of death equal to 1.26 for high-risk versus low-risk hospitals is clinically significant.

Panel Member #7: I have no concerns. A median odds ratio of death equal to 1.26 for high-risk versus low-risk hospitals is clinically significant.

Panel Member #9: The measure is able to identify a relatively small number of hospitals with either above-average performance or below-average performance. It does not seem to be capable of making any finer distinctions.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
Panel Member #1: N/A
Panel Member #3: N/A
Panel Member #5: NA
Panel Member #6: This domain is not applicable.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: No missing data in the development and testing data. No evaluation as to whether it would affect the hospital scores.

Panel Member #3: N/A, all elements were derived from claims and can be regarded to have no missing data.

Panel Member #4: No concerns

Panel Member #5: NA
Panel Member #6: This domain is not applicable.
Panel member #7: None.

16. **Risk Adjustment**

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

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d. **Risk adjustment summary:**

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

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

Panel Member #1: N/A

Panel Member #5: NA – factors present at start of care

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

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

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

16e**. Assess the risk-adjustment approach**

Panel Member #1: No issues.

Panel Member #4: The developers used a **Statistical risk model with 41 risk factors. They state: "**Our goal in selecting risk factors for adjustment was to develop parsimonious models that included clinically relevant variables strongly associated with the risk of mortality in the 30 days following an index admission. We used a two stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of social risk factors.." Further, they state "we find that the addition of any of these [SES] variables into the hierarchical model has little to no effect on hospital performance. And…" Overall, we find that among the SRF variables that could be feasibly incorporated into this model, the relationship between dual-eligible status and AHRQ low SES is in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also find that the impact of any of these indicators is small to negligible on model performance and hospital-level results. Given the controversial nature of incorporating such variables into a risk-model, we do not support doing so in a case that is unlikely to affect hospital profiling. Given these empiric findings, ASPE's recommendation to not risk adjust publicly reported quality measures for SRFs, and complex pathways which could explain the relationship between SRFs and mortality (and do not all support risk-adjustment), CMS chose to not incorporate SRF variables in this measure." In my opinion, this investigation is well done and I have no concerns.

Panel Member #5: The risk adjustment approach is sound and continues the track record of appropriately developing a risk model since the onset of the CMS 30 day mortality measures were created. Panel Member #6: The measure steward presents a comprehensive analysis of including social risk factors, including dual eligibility for Medicare and Medicaid and low socioeconomic status (as a function of 9-digit ZIP code) in the risk adjustment model. The inclusion of these variables does not alter model discrimination. More importantly, RMSR values for hospitals are almost perfectly correlated in risk adjustment approaches excluding and including these two social risk factors.

Panel Member #7: The developers provided substantial information on the risk factor model, including social factors, which they thoroughly discussed**.** They found that the C-Statistic (.73) did not change for the social risk factors considered (AHRQ SES variable based on 9-digit zip code and dual eligibility) , nor did the median change in hospital COPD RSMR and therefore did not adjust for social risk factors.

Panel Member #9: The developers have done an exemplary job of developing the risk-adjustment approach in general, and specifically in considering and analyzing the potential value of a set of social risk factors.

The analyses show that although several social risk factors could conceivably affect the outcome of COPD mortality, they in fact do not have a significant influence, and when included in a risk-adjustment model, do not have any meaningful effect. The decision to leave social risk factors out of the model is justified by the data analyses presented.

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

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

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

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

    Submission document: Testing attachment, section 2b2.2.

    Panel Member #1: Appropriate.

    Panel Member #3 Note: The NQF and SMP encourage developers to perform analyses comparing results of their measure to other related measures. For an outcome of obvious importance such as mortality, I don't attach high importance to a measure's correlation with other measures, especially if other available measures assess different domains or populations or are less relevant to patients and providers. My assessment of validity is largely based on whether I am convinced that estimates are relatively free from bias from factors such as case mix variation, data accuracy, etc.

    Panel Member #4: Empirical validity: The developer identified and assessed the measure's correlation with other measures that target the same domain of quality (e.g. complications, safety, or post-procedure utilization) for the same or similar populations (Hospital Star Rating mortality group score, Overall Hospital Star Rating). In addition, the developer assessed the Validity of Claims-Based Measures compared to medical records and assessed the Validity Indicated by Established Measure Development Guidelines.

    These methods seem appropriate to me. Face Validity as Determined by TEP:

    This is acceptable to me because empirical validity is also included.

    Panel Member #5: The empirical testing is appropriate for the measure. Meanwhile, face validity is only acceptable in light of also using empirical testing given this is a measure previously endorsed and in maintenance.

    Empirical Validity

- assessed the measure's correlation with other measures that target the same domain of quality:
- Hospital Star Rating mortality group score
- Overall Hospital Star Rating' [p12-13]
- TEP included 11 members including individuals with diverse perspectives and backgrounds,

  including clinicians, consumers, hospitals, purchasers, and experts in quality improvement' [p14]

Panel Member #6: Correlations of the measure with hospital star ratings mortality group scores and overall hospital star ratings were estimated.

Panel Member #7: The developer used two external validation variables to assess the COPD RSMR, the CMS Overall Hospital Star Rating and the Star Rating for the mortality group.

Panel Member #9: The face validity data carried over from the earlier round of endorsement review uses a strong method; the empirical analysis to establish validity is not convincing, as it involves two analyses of relationships – one with the Star Rating mortality scores, and one with the Star Rating overall scores. The first analysis shows a marginally significant relationship, but this is not impressive, as COPD mortality is one element of overall mortality, so some level of relationship is automatic as one measure is a component part of the other. The second analysis shows essentially no relationship, but no strong theoretical relationship was expected

22. **Assess the results(s) for establishing validity Submission document: Testing attachment, section2b2.3, 2b2.3**

Panel Member #1: Fair interpretation from developer.

Panel Member #3: No concerns, very rigorous.

Panel Member #4: Comparison to Star-Rating Mortality Scores

The developers state: "The correlation between COPD RSMRs and Star-Rating mortality score is -0.618, which suggests that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating mortality scores."

Comparison to Overall Star-Rating Scores

The developers state: "The correlation between COPD RSMRs and Star-Rating summary score is -0.165, which suggests that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating summary scores."

These results make sense to me and are acceptable.

Panel Member #5: The empirical testing result is moderate to strong regarding "comparison to star rating mortality scores", and weak to modest regarding "comparison to overall star rating, but in the appropriate direction.

Comparison to Star-Rating Mortality Scores

Figure 1 shows the box-whisker plots…The correlation between COPD RSMRs and Star-Rating mortality score is -0.618, which suggests that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating mortality scores.

Comparison to Overall Star-Rating Scores

Figure 2 shows the Box-whisker plots of the COPD mortality measure RSMRs within each quartile of Star-Rating summary scores. The blue circles represent the mean RSMRs of Star-Rating summary score quartiles. The correlation between COPD RSMRs and Star-Rating summary score is -0.165' [p16]Panel Member #6: Correlations of the measure with hospital star ratings mortality group scores and overall hospital star ratings were estimated.

Panel Member #9: Face validity was and is strong – results of empirical validity testing are weak.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☒ **No**

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

Panel Member #9: The appropriate answer here would be "marginally" – not clear either no or yes.

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

NOTE that data element validation from the literature is acceptable.

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

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

Panel Member #1: Validity testing results.

Panel Member #3: My rating is based on the fact that the methodological approach appears to be excellent and they rigorously address various concerns related to the selection of model covariates, model performance, etc.

Panel Member #4: No additional concerns.

Panel Member #5: Q22: The empirical testing result is moderate to strong regarding "comparison to star rating mortality scores", and weak to modest regarding "comparison to overall star rating, but in the appropriate direction.

Panel Member #6: The strength and directionality of the aforementioned correlations provides good evidence of validity.

Panel Member #7: Meaningful between facility differences are difficult to assess given the data provided or I would have rated the validation evidence as high. Panel Member #9: The moderate rating is due entirely to the strong face validity evidence presented at the earlier review round. The developers have done a fair job of at least trying to establish empirical validity, but the new evidence for empirical validity is not convincing at all. I stayed at moderate, though, because the new empirical validity evidence does not argue AGAINST the validity of the measure

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

Panel Member #5: No Concerns

Panel Member #9: It is disappointing that the measure developers could not find and present any more compelling evidence of measure validity. The measure therefore has marginal reliability, only the ability to identify a small percentage of above-average or below-average hospitals, and only strong face validity from a large panel of experts to support its use. There is no new empirical evidence to support the concept that this measure is a meaningful measure of hospital quality of care.

Developer Submission

NQF #: 1893

Corresponding Measures:

De.2. Measure Title: Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services

De.3. Brief Description of Measure: The measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR), defined as death from any cause within 30 days after the index admission date, for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD. CMS annually reports the measure for patients who are 65 years or older and enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals or are patients hospitalized in Veterans Health Administration (VA) facilities.

1b.1. Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for COPD. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Additionally, COPD mortality is a priority area for outcomes measure development, as it is a costly and common condition. Hospital mortality is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices. Furthermore, the measure will increase transparency for consumers and potentially lower the healthcare costs associated with mortality.

S.4. Numerator Statement: The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days from the date of admission for patients hospitalized with either a principal diagnosis of COPD or a principal diagnosis of respiratory failure with a secondary diagnosis of acute exacerbation of COPD.

S.6. Denominator Statement: This claims-based measure is used for a cohort of patients aged 65 years or older.

The cohort includes admissions for patients aged 65 years and older discharged from the hospital with a principal discharge diagnosis of COPD and with a complete claims history for the 12 months prior to admission. The measure is publicly reported by CMS for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details.

S.8. Denominator Exclusions: The mortality measures exclude index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;

2. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
3. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort for each year.

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Enrollment Data, Other

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jan 08, 2013 **Most Recent Endorsement Date:** Aug 03, 2016

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is paired with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization.

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

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

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

NQF_evidence_COPDmortality_Fall2020_final_7.22.20.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1893

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

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

Date of Submission: **11/2/2020**

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

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

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

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

Outcome

☒ Outcome: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

☐ Patient-reported outcome (PRO):

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

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

☐ Process:

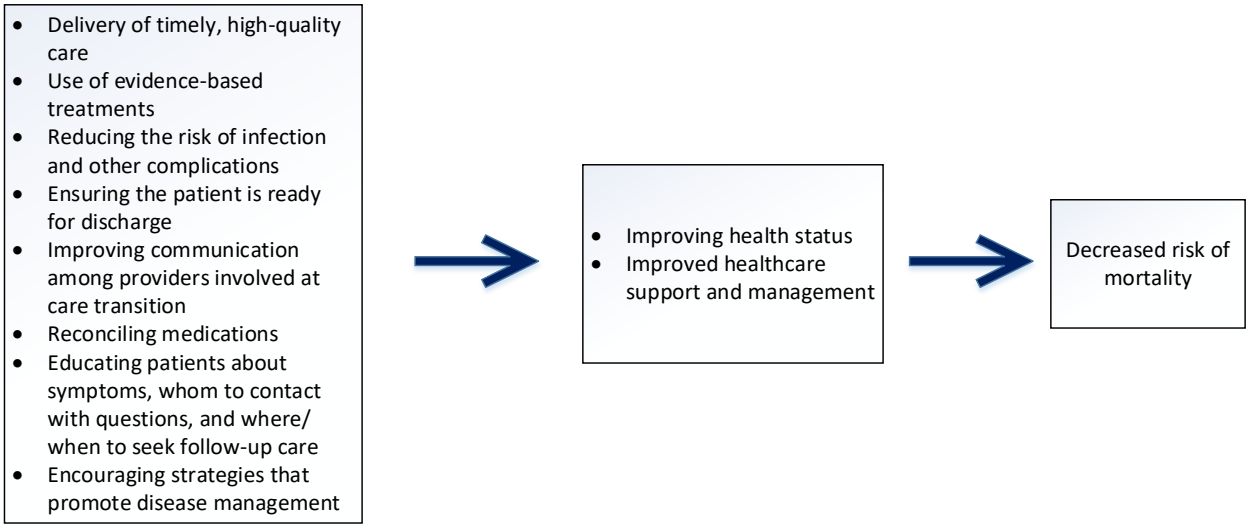
☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

Figure 1. COPD Mortality Logic Model



The goal of this measure is to directly affect patient outcomes by measuring risk-standardized rates of mortality. Measurement of patient outcomes, including mortality, allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. As described below, mortality is likely to be influenced by a broad range of clinical activities such as the prevention of complications and the provision of evidenced-based care.

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

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

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

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

COPD is a priority condition for outcomes measure development because it is a leading cause of morbidity and mortality. Although overall COPD prevalence has declined over the last decade (Biener et al., 2019), COPD continues to affect tens of millions of individuals in the United States and is the nation's fourth leading cause of death (National Heart, Lung, and Blood Institute, 2020; CDC, 2020; Wier et al., 2011; CDC, 2013; American Lung Association, 2015). Studies report that in-hospital mortality rates for patients hospitalized for exacerbations of COPD range from 2-5% (AHRQ, National Statistics on All Stays; Patil et al., 2009; Tabak et al.,

2009; Lindenauer et al., 2006; Dransfield et al., 2008) and 30-day mortality rates range from 3-9% (Faustini et al., 2008; Fruchter et al., 2008; Lindenauer et al., 2013). 30-day mortality rates following COPD discharge are also high and variable across hospitals; for the time period of July 2015-June 2018, publicly reported 30-day risk-standardized mortality rates among Medicare FFS patients ranged from 4.9% to 14.3% for patients admitted with COPD (Wallace et al., 2019).

In 2011 COPD was one of the top 20 most expensive conditions treated in U.S. hospitals (AHRQ, 2011). It was also one of the top 20 most expensive conditions billed to Medicare, accounting for nearly $4,074,000 of total hospital charges billed to Medicare (AHRQ, 2011). Some estimated project total costs of COPD treatment to increase to almost $50 billion dollars by 2020, primarily driven by disease complexity, lengthy hospital admissions, and increased prevalence of comorbid conditions (Lin et al., 2020; Guarascio et al., 2013; Huber et al., 2015).

Many current hospital processes have been associated with lower mortality rates within 30 days of hospital admission (Jha et al., 2007; Wright et al., 2019). In COPD in particular, supplemental oxygen and the use of noninvasive ventilation in carefully selected patients has been shown to improve both short- and long-term survival. Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. Measurement of patient outcomes, such as mortality, allows for a comprehensive view of quality of care that reflects complex aspects of care such as communication between providers and coordinated transitions to the outpatient environment. These aspects are critical to patient outcomes, and are broader than what can be captured by individual process-of-care measures.

The diagram above indicates some of the many care processes that can influence mortality risk. Numerous studies have demonstrated that appropriate (guideline recommended care) and timely treatment for COPD patients can reduce the risk of mortality within 30 days of hospital admission (Krumholz et al., 2007; Williams et al., 2012; Ram et al., 2004; Austin et al., 2010; Ntoumenopoulos, 2011). For instance, Wake Forest Baptist Medical Center showed promising reductions in readmissions and mortality after implementing a comprehensive care plan focused on transitions of care, treatment of comorbidities, and appropriate and timely hospice and palliative care services for COPD patients (Ohar et al., 2018).

The COPD mortality measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals (Krumholz et al., 2007).

References:

Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf. Accessed November 16, 2015.

Agency for Healthcare Research and Quality. Healthcare Cost and Utilization Project (HCUP). National Statistics on All Stays. Available at: http://hcupnet.ahrq.gov./HCUPnet.jsp?Id=DDB2B43489A1ADB8&Form=DispTab&JS=Y&Action=Accept. Accessed December 4, 2015.

American Lung Association. Lung Health & Diseases. Learn About Chronic Obstructive Pulmonary Disease (COPD)—How Serious is COPD, 2015. Available at: http://www.lung.org/lung-health-and-diseases/lung-disease-lookup/copd/learn-about-copd/how-serious-is-copd.html. Accessed November 16, 2015.

Austin MA, Wills KE, Blizzard L, et al. Effect of high flow oxygen on mortality in chronic obstructive pulmonary disease patients in prehospital setting: randomised controlled trial. BMJ. 2010 Oct 18; 341:c5462. doi: 10.1136/bmj.c5462.

Biener AI, Decker SL, Rohde F. Prevalence and Treatment of Chronic Obstructive Pulmonary Disease (COPD) in the United States. *JAMA.* 2019;322(7):602. doi:10.1001/jama.2019.10241.

Centers for Disease Control and Prevention. Detailed Tables for the National Vital Statistics Report (NVSR) "Deaths: Final Data for 2013". http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf Accessed November 16, 2015.

Centers for Disease Control and Prevention. Deaths: Final Data for 2017, Supplemental Tables. Chronic Obstructive Pulmonary Disease Among Adults — United States, 2020. Available at: https://www.cdc.gov/nchs/fastats/copd.htm. Accessed June 19. 2020.

Dransfield MT, Rowe SM, Johnson JE, et al. Use of beta blockers and the risk of death in hospitalized patients with acute exacerbations of COPD. Thorax. Apr 2008; 63 (4): 301-305.

Faustini A, Marino C, D'Ippoliti D, et al. The Impact on risk-factor analysis of different mortality outcomes in COPD patients. European Respiratory Journal. Sep 2008; 32 (3): 629-636.

Fruchter O, Yigla M., Predictors of long-term survival in elderly patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Respirology. Nov 2008; 13 (6): 851-855.

Guarascio AJ, Ray SM, Finch CK, Self TH. The clinical and economic burden of chronic obstructive pulmonary disease in the USA. *Clinicoecon Outcomes Res.* 2013;5:235-245. Published 2013 Jun 17. doi:10.2147/CEOR.S34321.

Huber MB, Wacker ME, Vogelmeier CF, Leidl R. Excess costs of comorbidities in chronic obstructive pulmonary disease: a systematic review. *PLoS One*. 2015;10(4):e0123292. Published 2015 Apr 13. doi:10.1371/journal.pone.0123292.

Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. Health Aff (Millwood) 2007 Jul-Aug; 26(4):1104-10.

Krumholz HM, Normand SL, Spertus JA, et al. Measuring performance for treating heart attacks and heart failure: the case for outcomes measurement. Health Aff (Millwood). 2007 Jan-Feb;26(1):75-85.

Lin SY, Xue H, Deng Y, Chukmaitov A. Multi-morbidities are Not a Driving Factor for an Increase of COPD-Related 30-Day Readmission Risk. *Int J Chron Obstruct Pulmon Dis*. 2020;15:143-154. Published 2020 Jan 15. doi:10.2147/COPD.S230072.

Lindenauer PK, Grosso LM, Wang C, et al. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. J Hosp Med. 2013 Aug; 8(8):428-35. doi: 10.1002/jhm.2066. Epub 2013 Jul 26.

Lindenauer PK, Pekow P, Gao S, et al. Quality of care for patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Ann Intern Med. June 20 2006; 144 (12): 894-903.

National Heart, Lung, and Blood Institute. What is COPD. 2020. Available at: https://www.nhlbi.nih.gov/health-topics/copd. Accessed June 16, 2020.

Ntoumenopoulos G. Using titrated oxygen instead of high flow oxygen during an acute exacerbation of chronic obstructive pulmonary disease (COPD) saves lives. J Physiother. 2011; 57(1):55. doi: 10.1016/S1836-9553(11)70008-X.

Ohar JA, Loh CH, Lenoir KM, Wells BJ, Peters SP. A comprehensive care plan that reduces readmissions after acute exacerbations of COPD. *Respir Med.* 2018;141:20-25.

Patil SP, Krishnan JA, Lechtzin n, Diette GB. In-hospital mortality following acute exacerbations of chronic obstructive pulmonary disease. Archives of Internal Medicine. Sep 28 2009; 169 (17): 1595-1602.

Ram FS, Picot J, Lightowler, Wedzicha JA. Non-invasive positive pressure ventilation for treatment of respiratory failure due to exacerbations of chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2004; (1):CD004104.

Tabak YP, Sun X, Johannes RS, et al. Mortality and need for mechanical ventilation in acute exacerbations of chronic obstructive pulmonary disease: development and validation of simple risk score. Archives of Internal Medicine. Sep 28 2009; 169(17): 1595-1602.

Wallace L., Grady J., Djordjevic, D., et al. 2019 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measures: Acute Myocardial Infarction – Version 13.0, Chronic Obstructive Pulmonary Disease – Version 8.0, Heart Failure – Version 13.0, Pneumonia – Version 13.0. 2019, Stroke – Version 8.0; https://qualitynet.org/inpatient/measures/mortality/methodology. Available as of April 23, 2019.

Wier LM, Elixhauser A, Pfuntner, et al. Overview of Hospitalizations among Patients with COPD, 2008. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs #106 [Internet]. Rockville (MD): Agency for Health Care Policy and Research (US). Feb. 2011. Available at: http://www.ncbi.nlm.nih.gov/books/NBK53969/. Accessed November 16, 2015.

Williams JW, Cox CE, Hargett CW, et al. Noninvasive Positive-Pressure Ventilation (NPPV) for Acute Respiratory Failure [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Jul. Report No.: 12-EHC089-EF. AHRQ Comparative Effectiveness Reviews.

Wright EA, Graham JH, Maeng D, et al. Reductions in 30-day readmission, mortality, and costs with inpatient-to-community pharmacist follow-up. *J Am Pharm Assoc (2003).* 2019;59(2):178-186.

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	*
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	*
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	*
Estimates of benefit and consistency across studies	*
What harms were identified?	*
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	*

*cell intentionally left blank

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

N/A

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

N/A

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

N/A

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

N/A

1b. Performance Gap

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

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

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

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

The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for COPD. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Additionally, COPD mortality is a priority area for outcomes measure development, as it is a costly and common condition. Hospital mortality is an outcome that is likely attributable to care processes and is an important outcome for patients. Measuring and reporting mortality rates will inform healthcare providers about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by Medicare patients. The measure will also provide patients with information that could guide their choices. Furthermore, the measure will increase transparency for consumers and potentially lower the healthcare costs associated with mortality.

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

Variation in mortality rates indicates opportunity for improvement. We conducted analyses using data from July 1, 2016 to June 30, 2019 Medicare claims and VA administrative data (n= 716,323 admissions from 4,642 hospitals).

The three-year hospital-level risk-standardized mortality rates (RSMRs) have a mean of 8.4% and range from 5.1-13.6% in the study cohort. As shown below, the median risk-standardized rate is 8.3%. The distribution of RSMRs across hospitals is shown below:

Distribution of Hospital COPD RSMRs over Different Time Periods

Results for each data year

Periods//YEAR1617//YEAR1718//YEAR1819//YEAR1619

Characteristic//07/2016-06/2017//07/2017-06/2018//07/2018-06/2019//07/2016-06/2019

Number of Hospitals//4,527//4,527//4,461//4,642

Number of Admissions//278,028//239,571//198,724//716,323

Mean(SD)//8.7(0.8)//8.4(0.6)//8(0.7)//8.4(1)

Range(Min-Max)// 5.6-13.9//5.5-11.9//5.5-13.8//5.1-13.6

Minimum//5.6//5.5//5.5//5.1

10th percentile//7.8//7.7//7.2//7.3

20th percentile//8.2//8.0//7.5//7.7

30th percentile//8.4//8.1//7.7//8.0

40th percentile//8.5//8.2//7.8//8.2

50th percentile//8.6//8.3//7.9//8.3

60th percentile//8.8//8.4//8.0//8.6

70th percentile//9.0//8.6//8.3//8.8

80th percentile//9.2//8.8//8.5//9.2

90th percentile//9.7//9.2//8.9//9.8

Maximum//13.9//11.9//13.8//13.6

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

N/A

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

Distribution of COPD RSMRs by Proportion of Dual Eligible Patients:

Data Source: Medicare FFS claims, VA data, and Medicare Beneficiary Summary File (MBSF) data

Dates of Data: July 2016 through June 2019

Variation in RSMRs across hospitals (with at least 25 cases) by proportion of patients with social risk//

Description of Social Risk Variable//Dual Eligibility

Quartile//Q1//Q4

Social Risk Proportion (%)//(0-9.76)//(36.49-100)

# of Hospitals//926//925

100%Max//13.0//13.4

90%//9.7//9.8

75%//9.0//9.0

50%//8.3//8.3

25%//7.7//7.7

10%//7.2//7.1

0%Min//5.1//5.1

Distribution of COPD RSMRs by Proportion of patients with AHRQ SES Index Scores:

Data Source: Medicare FFS claims, VA data, and the American Community Survey (ACS) data

Dates of Data: July 2016 through June 2019 (claims); 2013-2017 (ACS)

Variation in RSMRs across hospitals (with at least 25 cases) by proportion of patients in lower and upper social risk quartiles//

Description of Social Risk Variable //AHRQ SES Index

Quartile//Q1//Q4

Social Risk Proportion (%)//(0-12.42)//(27.72-98.46)

# of Hospitals//925//925

100%Max//13.4//13.4

90%//9.8//9.7

75%//9.0//8.9

50%//8.3//8.2

25%//7.7//7.6

10%//7.0//7.1

0%Min//5.6//5.1

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

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Respiratory, Respiratory : Chronic Obstructive Pulmonary Disease (COPD), Respiratory : Dyspnea

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

Care Coordination, Safety, Safety : Complications, Safety : Healthcare Associated Infections

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

Elderly, Populations at Risk

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

https://qualitynet.org/inpatient/measures/mortality/methodology

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

This is not an eMeasure **Attachment:**

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

Attachment: NQF_datadictionary_COPDmortality_Fall2020_final_7.22.20.xlsx

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

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

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

Not an instrument-based measure

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

No

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

Updates consisted of updating the specifications to include new and modified ICD-10 CM/PCS codes.

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

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

The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days from the date of admission for patients hospitalized with either a principal diagnosis of COPD or a principal diagnosis of respiratory failure with a secondary diagnosis of acute exacerbation of COPD.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

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

The measure counts all deaths (including in-hospital deaths) for any cause within 30 days of the date of the index COPD admission.

Identifying deaths in the FFS measure

As currently reported, we identify deaths for FFS Medicare patients 65 years and older in the Medicare Enrollment Database (EDB) and for VA patients in the VA data.

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

This claims-based measure is used for a cohort of patients aged 65 years or older.

The cohort includes admissions for patients aged 65 years and older discharged from the hospital with a principal discharge diagnosis of COPD and with a complete claims history for the 12 months prior to admission. The measure is publicly reported by CMS for those patients 65 years and older who are Medicare FFS or VA beneficiaries admitted to non-federal or VA hospitals, respectively.

Additional details are provided in S.7 Denominator Details.

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

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

To be included in the measure cohort used in public reporting, patients must meet the following inclusion criteria:

1. Principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD
2. Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the index admission and Part A during the index admission, or those who are VA beneficiaries
3. Aged 65 or over
4. Not transferred from another acute care facility.

This measure can also be used for an all-payer population aged 40 years and older. We have explicitly tested the measure in both patients aged 40+ years and those aged 65+ years (see Testing Attachment for details).

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

The mortality measures exclude index admissions for patients:

1. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
2. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
3. Discharged against medical advice (AMA).

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort for each year.

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

1. Inconsistent vital status or unreliable data are identified if any of the following conditions are met
    1) the patient's age is greater than 115 years:
    2) if the discharge date for a hospitalization is before the admission date;
    3) if the patient has a sex other than 'male' or 'female'.

Rationale: Reliable and consistent data are necessary for valid calculation of the measure.

2. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice data.

Rationale: These patients are likely continuing to seek comfort measures only; thus, mortality is not necessarily an adverse outcome or signal of poor quality care.

3. Discharges against medical advice (AMA) are identified using the discharge disposition indicator in claims data.

Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge.

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

N/A

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

The measure estimates hospital-level 30-day all-cause RSMRs following hospitalization for COPD using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of index admission using age, sex, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected mortality rates or better quality, and a higher ratio indicates higher-than-expected mortality rates or worse quality.

The "predicted" number of deaths (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of mortality. The estimated hospital-specific intercept is added coefficients multiplied by the patient characteristics. The results are transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of deaths (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are transformed and summed over all

patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report posted on QualityNet:
https://qualitynet.org/inpatient/measures/mortality/methodology.

References:

1. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

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

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

N/A. This measure is not based on a sample or survey.

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

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

N/A. This measure is not based on a sample or survey.

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

If other, please describe in S.18.

Claims, Enrollment Data, Other

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

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

Data sources for the Medicare FFS measure:

Medicare Part A Inpatient and Part B Outpatient Claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission.

Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992). The Master Beneficiary Summary File (MBSF) is an annually created file derived the EDB that contains enrollment information for all Medicare beneficiaries including dual eligible status. Years 2016-2019 were used.

Veterans Health Administration (VA) Data: This data source contains data for VA inpatient and outpatient services including: inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician data for the 12 months prior to and including each index admission. Unlike Medicare FFS patients, VA patients are not required to have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission.

The American Community Survey (2013-2017): The American Community Survey data is collected annually and an aggregated 5-years data were used to calculate the Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) composite index score.

References:

Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs hospitals. Medical Care. 1992; 30(5): 377-91.

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

No data collection instrument provided

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

Facility

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

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

NQF_testing_COPDmortality_Fall2020_final_10.27.20.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

Yes - Updated information is included

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

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

Measure Number (*if previously endorsed*)**:** 1893

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization

Date of Submission: **8/3/2020**

Type of Measure:

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

*cell intentionally left blank

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

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

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

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☒ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Medicare Enrollment Data, VHA Administrative Data	☒ other: Census Data/American Community Survey, VHA Administrative Data, Master Beneficiary Summary File

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

The data used for testing included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). Additionally, census as well as enrollment data were used to assess socioeconomic factors (dual eligible variable obtained through enrollment data; Agency for Healthcare Research and Quality [AHRQ] socioeconomic status [SES] index obtained through census data). Veterans' Health Administration (VHA) data are also included in the testing dataset. The dataset used varies by testing type; see Section 1.7 for details.

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

The dates used vary by testing type; see Section 1.7 for details.

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

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

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

For this measure, hospitals are the measured entities. All non-federal, short-term acute care inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years or over are included. In addition, for the testing period presented, VA hospitals and their patients 65 years and older are included in the measure. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

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

The number of admissions/patients varies by testing type: see Section 1.7 for details.

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

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are in Table 1.

Measure Development

For measure development, we used Medicare administrative claims data (2008). The dataset also included administrative data on each patient for the 12 months prior to the index admission. The dataset contained inpatient and facility outpatient claims and Medicare enrollment database (EDB) data. We randomly split the data into two equal samples: the Development Dataset and Internal Validation Dataset.

Measure Testing

For analytical updates for this measure, we used three-years of Medicare administrative claims data (July 2016 – June 2019). The dataset also included administrative data on each patient for the 12 months prior to the

index admission. The dataset contained inpatient and facility outpatient claims and Medicare enrollment database (EDB) data. The dataset also included administrative data from the VHA as these hospitals are currently publicly reported for this measure.

Table 1. Dataset Descriptions

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
Development and Validation Datasets (Medicare Fee-For-Service Administrative Claims Data)	Section 2b3 Risk Adjustment/Stratification **2b3.6. Statistical Risk Model Discrimination Statistics** **2b3.7. Statistical Risk Model Calibration Statistics**	Entire Cohort: Dates of Data: 2008 Number of admissions = 299,681 Patient Descriptive Characteristics: Number of measured hospitals: 4,357 This cohort was randomly split for initial model testing. First half of split sample -Number of Admissions: 150,035 -Number of Measured Hospitals: 4,537 Second half of split sample -Number of Admissions: 149,646 -Number of Measured Hospitals: 4,535
Testing Dataset (Medicare Fee-For-Service Administrative Claims Data (July 1, 2016 – June 30, 2019)	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b2 Testing of Measure Exclusion Section 2b3 Risk Adjustment/Stratification **2b3.6. Statistical Risk Model Discrimination Statistics** Section 2b4 Meaningful Differences	Dates of Data: July 2016 – June 2019 Number of admissions = 716,323 Patient Descriptive Characteristics: mean age = 76.8 years %Male = 44.5 Number of measured hospitals: 4,642 First half of split sample -Number of Admissions: 356,990 -Number of Measured Hospitals: 4,589 Mean age = 76.8 %Male = 44.5 Second half of split sample -Number of Admissions: 359,333 -Number of Measured Hospitals: 4,642 Mean age = 76.8

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
		%Male = 44.4
The American Community Survey (ACS)	Section 2b3: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: 2013-2017 We used the AHRQ SES index score derived from the American Community Survey (2013-2017) to study the association between the 30-day mortality outcome and SRFs. The AHRQ SES index score is based on beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in the American Community Survey.
Master Beneficiary Summary File (MBSF)	Section 2b3: Risk adjustment/Stratification for Outcome or Resource Use Measures	Dates of Data: July 2016 – June 2019 We used dual eligible status (for Medicare and Medicaid) derived from the MBSF to study the association between the 30-day measure outcome and dual-eligible status.

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

We selected social risk factor (SRF) variables to analyze after reviewing the literature and examining available national data sources. We sought to find variables that are consistently captured in a reliable fashion for all patients in this measure. There is a large body of literature linking various SRFs to worse health status and higher mortality over a lifetime. Income, education, and occupation are the most commonly examined SRFs studied. The causal pathways for SRF variable selection are described below in Section 2b3.3a. Unfortunately, these variables are not available at the patient level for this measure. Therefore, proxy measures of income, education level and economic status were selected.

The SRF variables used for analysis were:

- Dual eligible status: Dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data is obtained from the CMS Master Beneficiary Summary File (MBSF).

  Following guidance from ASPE and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable (ASPE, 2016; ASPE, 2020). We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous outcome. However, the threshold for over 65-year-old Medicare patients is valuable, as it takes into account both income and assets and is consistently applied across states for the older population. We acknowledge that it is important to test a wider variety of SRFs including key variables such as education and poverty level; therefore, we also tested a validated composite based on census data linked to as small a geographic unit as possible.

- AHRQ-validated SES index score (summarizing the information from the following 7 variables): percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room)

  Finally, we selected the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas (Bonito et al., 2008). Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of SDS indicators. We ultimately did not include the ADI at the time, partly due to the fact that the coefficients used to derive ADI had not been updated for many years. Recently, the coefficients for ADI have been updated and therefore we compared the ADI with the AHRQ SES Index and found them to be highly correlated. In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 9-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. We used the percentage of patients with an AHRQ SES index score equal to or below 42.7 to define the lowest quartile of the AHRQ SES Index.

References:

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health affairs (Project Hope). 2002; 21(2):60-76.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation. Cardiovascular quality and outcomes. May 2014; 7(3):391-397.

Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;2.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed November 10, 2019.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.

Eapen ZJ, McCoy LA, Fonarow GC, Yancy CW, Miranda ML, Peterson ED, Califf RM, Hernandez AF. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.

Gilman M, Adams EK, Hockenberry JM, Wilson IB, Milstein AS, Becker ER. California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Aff (Millwood). Aug 2014; 33(8):1314-22.

Glymour MM, Kosheleva A, Boden-Albala B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. Neurology. Dec 1 2009;73(22):1858-1865.

Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

Kosar CM, Loomer L, Ferdows NB, Trivedi AN, Panagiotou OA, Rahman M. Assessment of Rural-Urban Differences in Postacute Care Utilization and Outcomes Among Older US Adults. *JAMA Netw Open*. 2020;3(1):e1918738. Published 2020 Jan 3. doi:10.1001/jamanetworkopen.2019.18738.

Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. European heart journal. 2000; 21(14):1141-1151.

Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PloS one. 2011;6(7):e22693.

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. Circulation. Jun 14 2005; 111(23):3063-3070.

2a2. RELIABILITY TESTING

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

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

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

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

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

Measure Score Reliability

We performed two types of reliability testing. First, we estimated the overall measure score reliability by calculating the intra-class correlation coefficient (ICC) using a split sample (i.e. test-retest) method. Second, we estimated the facility-level reliability (signal-to-noise reliability).

Split-Sample Reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is naturally the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. Accordingly, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produce similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, and then measured again using a second random subset exclusive of the first, and the agreement of the two resulting performance measures compared across hospitals (Rousson, Gasser, and Seifert, 2002).

For split-sample reliability of the measure in aged 65 years and older, we randomly sampled half of patients within each hospital for a three-year period, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (Shrout & Fleiss, 1979), and assessed the values according to conventional standards (Landis & Koch, 1977). Specifically, we used a combined 2016-2019 sample, randomly split it into two approximately equal subsets of patients, and calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1) (Shrout & Fleiss, 1979).

Using two non-overlapping random samples provides a conservative estimate of the measure's reliability, compared with using two random but potentially overlapping samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that smaller volume hospitals contribute less 'signal', a split sample using a single measurement period would introduce extra noise. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman 1910, Brown 1910). We used this formula to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

Signal-to-Noise

We estimated the signal to noise reliability (facility-level reliability), which is the reliability with which individual units (hospitals) are measured. While split-sample reliability is the most relevant metric from the perspective of overall measure reliability, it is also meaningful to consider the separate notion of "unit" reliability, that is, the reliability with which individual units (here, hospitals) are measured. The reliability of any one facility's measure score will vary depending on the number of patients admitted for COPD. Facilities with more volume (i.e., with more patients) will tend to have more reliable scores, while facilities with less volume will tend to have less reliable scores. Therefore, we used the formula presented by Adams and colleagues (2010) to calculate facility-level reliability.

Where facility-to-facility variance is estimated from the hierarchical logistic regression model, n is equal to each facility's observed case size, and the facility error variance is estimated using the variance of the logistic distribution (π^2/3). The facility-level reliability testing is limited to facilities with at least 25 admissions for public reporting.

Signal to noise reliability scores can range from 0 to 1. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real difference in performance.

Additional Information

In constructing the measure, we aim to utilize only those data elements from the claims that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard.

In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

Furthermore, we assessed the variation in the frequency of the variables over time. Detailed information is presented in the measure's 2020 Condition-Specific Measure Updates and Specifications Report cited below.

References

Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

Debuhr J, McDowell K, Grady J et al., 2020 Condition-Specific Measure Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measures - Available at: https://www.qualitynet.org/inpatient/measures/mortality/methodology

Landis J, Koch G, The measurement of observer agreement for categorical data, Biometrics, 1977;33:159-174.

Rousson V, Gasser T, Seifert B. "Assessing intrarater, interrater and test–retest reliability of continuous measurements," Statistics in Medicine, 2002, 21:3431-3446.

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-3428.

Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

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

Measure Score Reliability Results

Split-Sample Reliability

In total, 716,323 admissions were included in the analysis, using three years of data. After randomly splitting the sample into two halves, there were 356,990 admissions from 4,589 hospitals in one half and 359,333 admissions from 4,642 hospitals in the other half. As a metric of agreement, we calculated the ICC for hospitals with 25 admissions or more. Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the RSMR for each hospital was 0.477.

Signal-to-Noise

We calculated the signal-to-noise reliability score for each hospital with at least 25 admissions* (see Table 2 below). The median reliability score was 0.72, ranging from 0.32 to 0.97. The 25th and 75th percentiles were 0.54 and 0.83, respectively. The median reliability score demonstrates moderate reliability.

Table 2

Mean	Std. Dev.	Min	5th Percentile	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	95th Percentile	Max
0.68	0.17	0.32	0.37	0.42	0.54	0.72	0.83	0.88	0.91	0.97

*Hospital measure scores are calculated for all hospitals (including those that have fewer than 25 admissions) but only publicly reported for those that have at least 25 admissions to ensure hospital results are reliable.

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

Measure Score Reliability Results

The split-sample reliability score of 0.477, discussed in the previous section, represents the lower bound of estimate of the true measure reliability.

Using the approach used by Adams et. al. and Yu et al., we obtained the median signal-to-noise reliability score of 0.72, which demonstrates substantial agreement.

Our interpretation of the results is based on the standards established by Landis and Koch (1977):

< 0 – Less than chance agreement;

0 – 0.2 Slight agreement;

0.21 – 0.39 Fair agreement;

0.4 – 0.59 Moderate agreement;

0.6 – 0.79 Substantial agreement;

0.8 – 0.99 Almost Perfect agreement; and

1 Perfect agreement

In the absence of empirically supported standards, our position is that 'acceptability' depends on context. For simple concepts or constructs, such as a patient's weight, the expectation is that the test-retest reliability of a measure of that construct should be quite high. However, for complex constructs, such as clinical severity, patient comorbidity, or symptom profiles used to identify a condition or clinical state, reliability of measures used to define these constructs is quite a bit lower.

Taken together, these results indicate that there is moderate reliability in the measure score.

References:

Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

Landis J, Koch G. The measurement of observer agreement for categorical data, Biometrics 1977;33:159-174.

Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Empirical Validity

Stewards of NQF-endorsed measures going through the re-endorsement process are required to demonstrate external validity testing at the time of maintenance review, or if this is not possible, justify the use of face validity only. To meet this requirement for the COPD mortality measure, we identified and assessed the measure's correlation with other measures that target the same domain of quality (e.g. complications, safety, or post-procedure utilization) for the same or similar populations. The goal was to identify if better performance in this measure was related to better performance on other relevant structural or outcomes measures. After literature review and consultations with measures experts in the field, there were very few measures identified that assess the same domains of quality. Given that challenge, we selected the following to use for validity testing.

> Hospital Star Rating mortality group score: CMS's Hospital Star Rating mortality group score assesses hospitals' overall performance (expressed on Hospital Compare graphically, as stars) based on a weighted average of group scores from the mortality domain. The mortality group is comprised of the mortality measures that are publicly reported on Hospital Compare, including this measure. The mortality group score is derived from a latent-variable model that identifies an underlying quality trait for that group. For the validity testing presented in this testing form, we used mortality group scores from 4,642 Medicare FFS hospitals from July 2019. The full methodology for the Overall Hospital Star Rating can be found at: https://www.qualitynet.org/inpatient/public-reporting/overall-ratings/resources.

Overall Hospital Star Rating: CMS's Overall Hospital Star Rating assesses hospitals' overall performance (expressed on Hospital Compare graphically, as stars) based on a weighted average of "group scores" from different domains of quality (mortality, readmissions, safety, patient experience, imaging, effectiveness of care, timeliness of care). Each group has within it, measures that are reported on Hospital Compare. Group scores for each individual group are derived from latent-variable models that identify an underlying quality trait for each group. Group scores are combined into an overall hospital score using fixed weights; overall hospital scores are then clustered, using k-means clustering, into five groups and are assigned one-to-five stars (the hospital's Star Rating). For the validity testing presented in this testing form, we used hospital's Star Ratings from 4,642 Medicare FFS hospitals from July 2019. The full methodology for the Overall Hospital Star Rating can be found at https://www.qualitynet.org/inpatient/public-reporting/overall-ratings/resources.

We examined the relationship of performance the COPD mortality measure scores (RSMR) with each of the external measures of hospital quality. For the external measures, the comparison was against performance within quartiles of the mortality group score, or in the case of Star Ratings, to the Star Rating category (1-5 Stars). We predicted the COPD mortality scores would be more strongly associated with the Hospital Star Rating mortality group score than the Overall Star Ratings scores, with lower RSMRs associated with better Star Ratings.

Validity of Claims-Based Measures

During measure development CORE validated the performance of the claims-based model and a medical records-based model and found the performance was similar. The areas under the receiver operating characteristic (ROC) curve are 0.69 and 0.77, respectively, for the two models. We estimated hospital-level RSMRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample. We then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital. The correlation coefficient of the standardized rates from the administrative and medical record models is 0.91 which shows that there was a strong correlation in rates calculated from the clinical and administrative models.

Our team has demonstrated for a number of prior measures the validity of claims-based measures for profiling hospitals by comparing either the measure results or individual data elements against medical records. CMS validated seven NQF-endorsed measures currently in public reporting (AMI, heart failure, and pneumonia mortality and readmission and coronary artery bypass graft surgery or CABG readmission) with models that used chart-abstracted data for risk-adjustment. Specifically, claims model validation was conducted by building comparable models using abstracted medical chart data for risk-adjustment for heart failure patients (National Heart Failure data) (Krumholz et al., 2006 [3]; Keenan et al., 2008), AMI patients (Cooperative Cardiovascular Project data) (Krumholz et al., 2006 [2]), pneumonia patients (National Pneumonia Project dataset) (Bratzler et al., 2011), and CABG patients (Shahian et al., 2014; Suter et al., 2014). When both models were applied to the same patient population, the hospital risk-standardized rates estimated using the claims-based risk adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based models for public reporting.

We have also completed two national, multi-site validation efforts for two procedure-based complications measures (for primary elective hip/knee arthroplasty and implantable cardioverter defibrillator [ICD]). Both projects demonstrated strong agreement between complications coded in claims and abstracted medical chart data.

Validity Indicated by Established Measure Development Guidelines:

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006 [1]).

Validity as Assessed by External Groups:

Throughout measure development, we obtained expert and stakeholder input via three mechanisms: regular discussions with an advisory working group, a national Technical Expert Panel (TEP), and a 30-day public comment period in order to increase transparency and to gain broader input into the measure.

The working group was assembled, and regular meetings were held throughout the development phase. The working group was tailored for development of this measure and consisted of three physicians who are board-certified in pulmonary and critical care medicine and a pharmacoepidemiologist with expertise in COPD. All members have expertise in quality measure development. The working group meetings addressed key issues related to measure development, including weighing the pros and cons of and finalizing key decisions (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The working group provided a forum for focused expert review and discussion of technical issues during measure development prior to consideration by the broader TEP.

In addition to the working group, and in alignment with the CMS MMS, we convened a TEP to provide input and feedback during measure development from a group of recognized experts in relevant fields. To convene the TEP, we released a public call for nominations and selected individuals to represent a range of perspectives, including physicians, consumers, and purchasers, as well as individuals with experience in quality improvement, performance measurement, and health care disparities. We held three structured TEP conference calls consisting of presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site: https://www.cms.gov/MMS/17_CallforPublicComment.asp. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and contributed to minor modifications to the measure.

Face Validity as Determined by TEP:

One means of confirming the validity of this measure was face validity assessed by our Technical Expert Panel (TEP), which included 11 members including individuals with diverse perspectives and backgrounds, including clinicians, consumers, hospitals, purchasers, and experts in quality improvement.

List of TEP Members

- Darlene Bainbridge, MS, NHA, CPHQ, CPHRM (President/CEO, Darlene D. Bainbridge & Associates, Inc.)
- Robert A. Balk, MD (Director of Pulmonary and Critical Care Medicine, Rush University Medical Center)
- Dale Bratzler, DO, MPH (President and CEO, Oklahoma Foundation for Medical Quality)
- Scott Cerreta, RRT (Director of Education, COPD Foundation)
- Gerard J. Criner, MD (Director of Temple Lung Center and Divisions of Pulmonary and Critical Care Medicine, Temple University)
- Guy D'Andrea, MBA (President, Discern Consulting)
- Jonathan Fine, MD (Director of Pulmonary Fellowship, Research and Medical Education, Norwalk Hospital)
- David Hopkins, MS, PhD (Senior Advisor, Pacific Business Group on Health)
- Fred Martin Jacobs, MD, JD, FACP, FCCP, FCLM (Executive Vice President and Director, Saint Barnabas Quality Institute)
- Natalie Napolitano, MPH, RRT-NPS (Respiratory Therapist, Inova Fairfax Hospital)
- Russell Robbins, MD, MBA (Principal and Senior Clinical Consultant, Mercer)

We systematically assessed the face validity of the measure score as an indicator of quality by soliciting the TEP members' agreement with the following statement: "The risk-standardized mortality rates obtained from the COPD mortality measure as specified will provide an accurate reflection of quality."

On a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5=Moderately Agree, and 6=Strongly Agree), 10 of 11 TEP members responded to the survey question as follows: Strongly Disagreed (1), Somewhat Agreed (3), Moderately Agreed (4), and Strongly Agreed (2). Of the TEP members who responded, 90% agreed (60% moderately or strongly agreed) that the measure will provide an accurate reflection of quality. We therefore gave the measure a moderate rating for face validity. In summary, these results demonstrated TEP agreement with the overall face validity of the measure as specified.

References:

Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. PLoS One 2011; 6(4):e17401.

Keenan PS, Normand SL, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation 2008; 1(1):29-37.

Krumholz HM, Brindis RG, Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation. January 24, 2006; 113(3):456-462. [1]

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. Circulation 2006; 113(13):1683-92. [2]

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation 2006; 113:1693-1701. [3]

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Shahian DM, He X, O'Brien S, et al. Development of a Clinical Registry-Based 30-Day Readmission Measure for Coronary Artery Bypass Grafting Surgery. Circulation 2014; DOI: 0.1161/CIRCULATIONAHA.113.007541. Published online before print June 10, 2014.

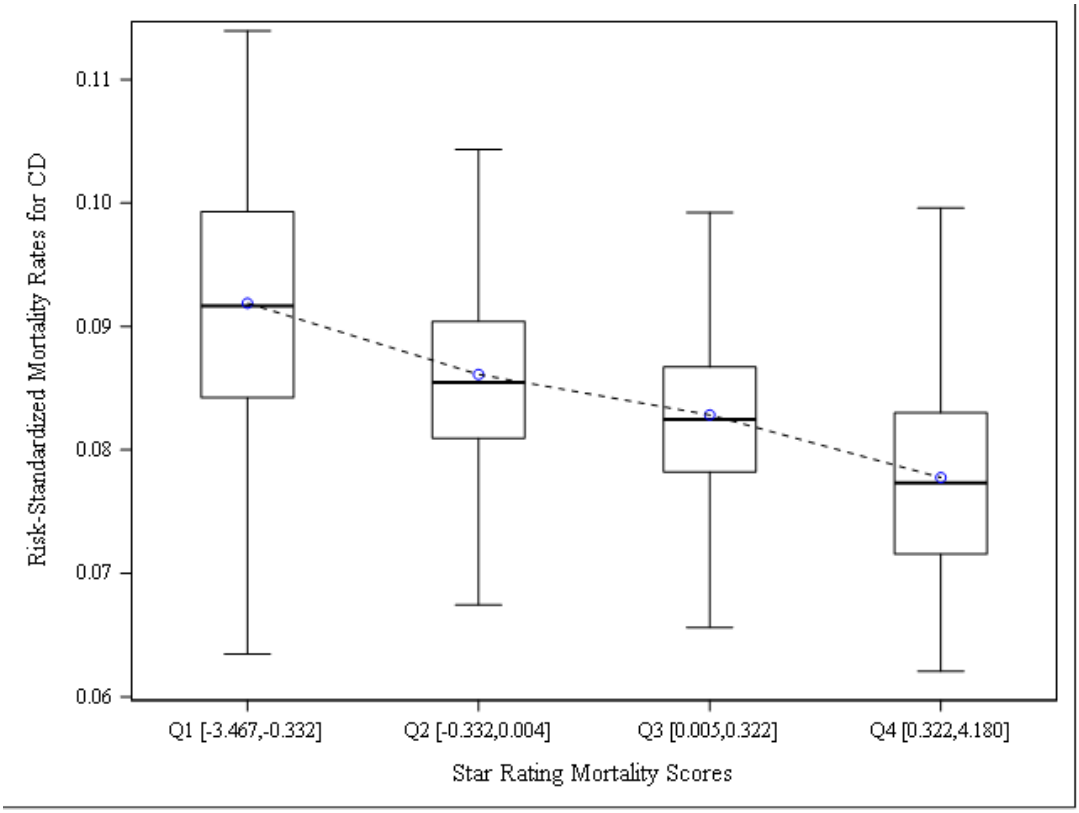
Suter L, Wang C, Araas M, et al. Hospital-Level 30-Day All-Cause Unplanned Readmission Following Coronary Artery Bypass Graft Surgery (CABG): Updated Measure Methodology Report. 2014; https://www.qualitynet.org/inpatient/measures/readmission/methodology. Accessed November 4, 2015.

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

Comparison to Star-Rating Mortality Scores

Figure 1 shows the box-whisker plots of the COPD mortality measure RSMRs within each quartile of Star-Rating mortality scores. The blue circles represent the mean RSMRs of Star-Rating mortality score quartiles. The correlation between COPD RSMRs and Star-Rating mortality score is -0.618, which suggests that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating mortality scores.

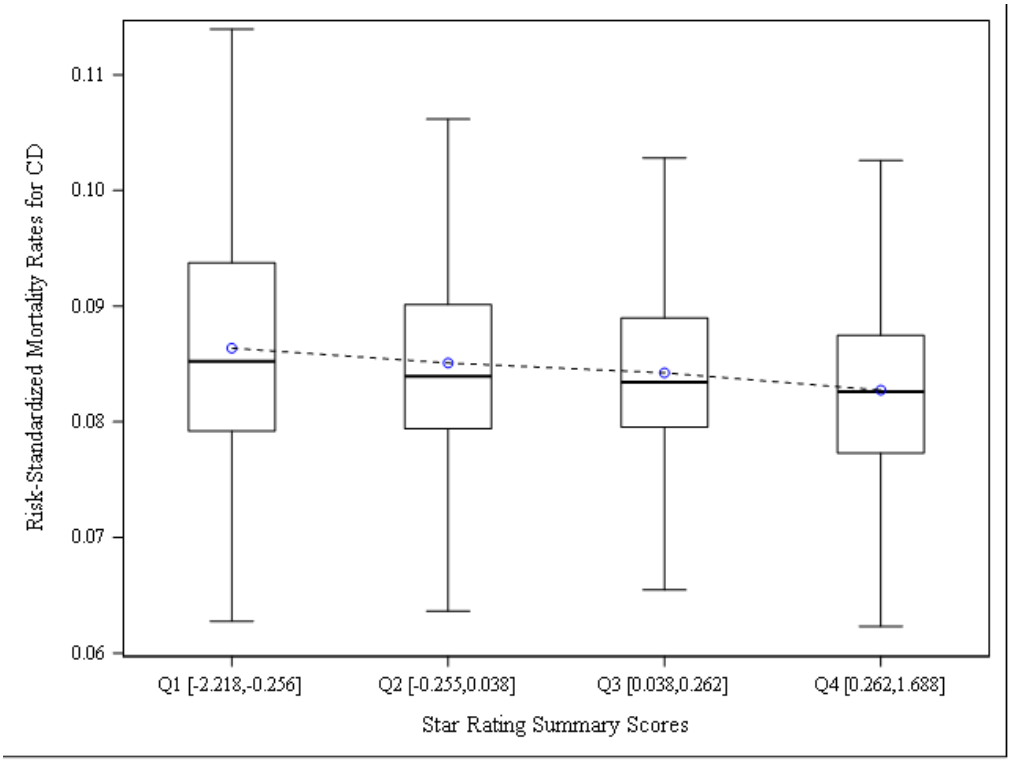
Figure 1



Comparison to Overall Star-Rating Scores

Figure 2 shows the Box-whisker plots of the COPD mortality measure RSMRs within each quartile of Star-Rating summary scores. The blue circles represent the mean RSMRs of Star-Rating summary score quartiles. The correlation between COPD RSMRs and Star-Rating summary score is -0.165, which suggests that hospitals with lower COPD RSMRs are more likely to have higher Star-Rating summary scores.

Figure 2



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

Empirical Validity Testing

This validation approach compares the 30-day COPD mortality measure results against the star rating mortality domain and overall summary scores. Figure 1 and 2 Box Plots results demonstrate an observed trend of lower risk-standardized mortality with higher star ratings score, especially at the extremes, which supports measure score validity. The correlation coefficients associated with the star rating mortality domain scores and the COPD mortality measure scores indicate a strong association. A more moderate association is seen with the overall star ratings score, which is to be expected given the measures are calculated by complex statistical models. Overall, the results above show that the trend and direction of this association is in line with what would be expected.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

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

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Testing Dataset**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.9 (Denominator Exclusions).

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

In the **Testing Dataset (Table 3)**, below is the distribution of exclusions among hospitals with 25 or more admissions:

Exclusion	N	%	Distribution across hospitals (*N=3,850*): Min, 25th, 50th, 75th percentile, max
1. Inconsistent or unknown vital status or other unreliable demographic data	49	0.01	(0.00, 0.00, 0.00, 0.00, 2.78)
2. Enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission	14,763	1.58	(0.00, 0.00, 1.08, 2.22, 19.7)
3. Discharged against medical advice (AMA)	7,122	0.76	(0.00, 0.00, 0.41, 1.07, 13.8)

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

Exclusion 1 (patients with inconsistent or unknown vital status or other unreliable demographic [age and gender] data), we do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive. This exclusion accounts for <0.01% of all index admissions excluded from the initial index cohort.

Exclusion 2 (patients enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission), these patients are likely continuing to seek comfort measures only; mortality is not necessarily an adverse outcome or signal of poor quality care. This exclusion accounts for 1.58% of all index admissions excluded from the initial index cohort.

Exclusion 3 (patients who are discharged AMA) accounts for 0.76% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with a similar probability of the outcome. For each patient, the probability of death changes with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three-year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 41 risk factors**

☐ **Stratification by risk categories**

☐ **Other**

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

See risk model specifications in Section 2b3.4a and the attached data dictionary.

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

N/A. This measure is risk-adjusted.

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

Selecting Risk Variables

Our goal in selecting risk factors for adjustment was to develop parsimonious models that included clinically relevant variables strongly associated with the risk of mortality in the 30 days following an index admission. We used a two stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of social risk factors.

The original measure was developed with ICD-9. When ICD-10 became effective in 2015, we transitioned the measure to use ICD-10 codes as well. ICD-10 codes were identified using 2015 GEM mapping software. We then enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes used to define this measure during development. A code set is attached in field S.2b. (Data Dictionary).

For risk model development, we started with Condition Categories (CCs) which are part of CMS's Hierarchical Condition Categories (HCCs). The current HCC system groups the 70,000+ ICD-10-CM and 17,000+ ICD-9-CM codes into larger clinically coherent groups (201 CCs) that are used in models to predict mortality or other outcomes (Pope et al. 2001; 2011). The HCC system groups ICD- codes into larger groups that are used in models to predict medical care utilization, mortality, or other related measures.

To select candidate variables, a team of clinicians reviewed all CCs and excluded those that were not relevant to the Medicare population or that were not clinically relevant to the mortality outcome (for example, attention deficit disorder, female infertility). All potentially clinically relevant CCs were included as candidate variables and, consistent with CMS's other claims-based mortality measures, some of those CCs were then combined into clinically coherent CC groupings.

To inform final variable selection, a modified approach to stepwise logistic regression was performed. The Development Sample was used to create 1,000 "bootstrap" samples. For each sample, we ran a logistic stepwise regression that included the candidate variables. The results (not shown in this report) were summarized to show the percentage of times that each of the candidate variables was significantly associated with mortality ($p<0.01$) in each of the 1,000 repeated samples (for example, 90 percent would mean that the candidate variable was selected as significant at $p<0.01$ in 90 percent of the times). We also assessed the direction and magnitude of the regression coefficients.

The clinical team reviewed these results and decided to retain risk adjustment variables above a predetermined cutoff, because they demonstrated a strong and stable association with risk of mortality and were clinically relevant. Additionally, specific variables with particular clinical relevance to the risk of mortality were forced into the model (regardless of percent selection) to ensure appropriate risk adjustment for COPD. These included:

Markers for end of life/frailty:

- Cancers (CC 8-CC 9)
- Hemiplegia, Paraplegia, Paralysis, Functional disability (CC 70-CC 74, CC 103, CC 104, CC 189-CC 190)
- Stroke (CC 99-CC 100)
- Head injury (CC 166-168)
- Major fracture, except of skull, vertebrae, or hip (CC 171),
- Traumatic amputations and complications (CC 173)

This resulted in a final risk-adjustment model that included 41 variables.

Social Risk Factors

We weigh SRF adjustment using a comprehensive approach that evaluates the following:

- Well-supported conceptual model for influence of SRFs on measure outcome (detailed below);
- Feasibility of testing meaningful SRFs in available data (section 1.8); and
- Empiric testing of SRFs (section 2b3.4b).

Below, we summarize the findings of the literature review and conceptual pathways by which social risk factors may influence risk of the outcome, as well as the statistical methods for SRF empiric testing. Our conceptualization of the pathways by which patients' social risk factors affect the outcome is informed by the literature cited below and IMPACT Act–funded work by the National Academy of Science, Engineering and Medicine (NASEM) and the Department of Health and Human Services Assistant Secretary for Policy and Evaluation (ASPE).

Causal Pathways for Social Risk Variable Selection

Although some recent literature evaluates the relationship between patient SRFs and the mortality outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways (see, for example, Chang et al 2007; Gopaldas et al., 2009; Kim et al., 2007; LaPar et al., 2010; 2012; Lindenauer et al., 2013; Trivedi et al., 2014; Buntin et al., 2017; Kosar et al., 2020). Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with mortality.

The social risk factors that have been examined in the literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables.

Patient-level variables describe characteristics of individual patients, and include the patient's income or education level (Eapen et al., 2015). Neighborhood/community-level variables use information from sources such as the American Community Survey as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the AHRQ-validated SES index score (Blum et al., 2014). Some of these variables may include the local availability of clinical providers (Herrin et al., 2015; Herrin et al., 2016). Hospital-level variables measure attributes of the hospital which may be related to patient risk (Roshanghalb et al., 2019). Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital (Gilman et al., 2014; Jha et al., 2013).

The conceptual relationship, or potential causal pathways by which these possible social risk factors influence the risk of mortality following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider:

1. **Patients with social risk factors may have worse health at the time of hospital admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and

may present for their hospitalization or procedure with a greater severity of underlying illness. These social risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment.

2. **Patients with social risk factors often receive care at lower quality hospitals**. Patients of lower income, lower education, or unstable housing have inequitable access to high quality facilities, in part, because such facilities are less likely to be found in geographic areas with large populations of poor patients. Thus, patients with low income are more likely to be seen in lower quality hospitals, which can explain increased risk of mortality following hospitalization.
3. **Patients with social risk factors may receive differential care within a hospital**. The third major pathway by which social risk factors may contribute to mortality risk is that patients may not receive equivalent care within a facility. For example, patients with social risk factors such as lower education may require differentiated care (e.g. provision of lower literacy information – that they do not receive).
4. **Patients with social risk factors may experience worse health outcomes beyond the control of the health care system.** Some social risk factors, such as income or wealth, may affect the likelihood of mortality without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing financial priorities which don't allow for adequate recuperation or access to needed treatments, or a lack of access to care outside of the hospital.

Although we analytically aim to separate these pathways to the extent possible, we acknowledge that risk factors often act on multiple pathways, and as such, individual pathways are complex to distinguish analytically. Further, some social risk factors, despite having a strong conceptual relationship with worse outcomes, may not have statistically meaningful effects on the risk model. They also have different implications on the decision to risk adjust or not.

Based on this model and the considerations outlined in section 1.8 – namely, that the AHRQ SES index and dual eligibility variables aim to capture the SRFs that are likely to influence these pathways (income, education, housing, and community factors) - the following social risk variables were considered for risk-adjustment:

- Dual eligible status
- AHRQ SES index

References

Barnato AE, Lucas FL, Staiger D, Wennberg DE, Chandra A. Hospital-level Racial Disparities in Acute Myocardial Infarction Treatment and Outcomes. Medical care. 2005;43(4):308-319.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation Cardiovascular quality and outcomes 2014; 7:391-7.

Buntin MB, Ayanian JZ. Social Risk Factors and Equity in Medicare Payment. *New England Journal of Medicine.* 2017;376(6):507-510.

Calvillo-King L, Arnold D, Eubank KJ, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. J Gen Intern Med. 2013 Feb; 28(2):269-82. doi: 10.1007/s11606-012-2235-x. Epub 2012 Oct 6.

Chang W-C, Kaul P, Westerhout C M, Graham M. M., Armstrong Paul W., "Effects of Socioeconomic Status on Mortality after Acute Myocardial Infarction." The American Journal of Medicine. 2007; 120(1): 33-39.

Committee on Accounting for Socioeconomic Status in Medicare Payment Programs; Board on Population Health and Public Health Practice; Board on Health Care Services; Institute of Medicine; National Academies of Sciences, Engineering, and Medicine. Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors. Washington (DC): National Academies Press (US); 2016 Jan 12. (https://www.ncbi.nlm.nih.gov/books/NBK338754/doi:10.17226/21858

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation. Report to Congress: Social Risk Factors and Performance under Medicare's Value-based Payment Programs. December 21, 2016. (https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs.

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.

Eapen ZJ, McCoy LA, Fonarow GC, Yancy CW, Miranda ML, Peterson ED, Califf RM, Hernandez AF. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.

Gilman M, Adams EK, Hockenberry JM, et al. California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Aff (Millwood). Aug 2014; 33(8):1314-22.

Gopaldas R R, Chu D., "Predictors of surgical mortality and discharge status after coronary artery bypass grafting in patients 80 years and older." The American Journal of Surgery. 2009; 198(5): 633-638.

Hasnain-Wynia R, Kang R, Landrum MB, Vogeli C, Baker DW, Weissman JS. Racial and ethnic disparities within and between hospitals for inpatient quality of care: an examination of patient-level Hospital Quality Alliance measures. Journal of health care for the poor and underserved. May 2010;21(2):629-648.

Herrin J, Kenward K, Joshi MS, Audet AM, Hines SJ. Assessing Community Quality of Health Care. Health Serv Res. 2016 Feb;51(1):98-116. doi: 10.1111/1475-6773.12322. Epub 2015 Jun 11. PMID: 26096649; PMCID: PMC4722214.

Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health Serv Res. 2015 Feb;50(1):20-39. doi: 10.1111/1475-6773.12177. Epub 2014 Apr 9. PMID: 24712374; PMCID: PMC4319869.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs 2011; 30:1904-11.

Kim C, Diez A V, Diez Roux T, Hofer P, Nallamothu B K, Bernstein S J, Rogers M, "Area socioeconomic status and mortality after coronary artery bypass graft surgery: The role of hospital volume." Clinical Investigation Outcomes, Health Policy, and Managed Care. 2007; 154(2): 385-390.

Kosar CM, Loomer L, Ferdows NB, Trivedi AN, Panagiotou OA, Rahman M. Assessment of Rural-Urban Differences in Postacute Care Utilization and Outcomes Among Older US Adults. *JAMA Netw Open*. 2020;3(1):e1918738. Published 2020 Jan 3. doi:10.1001/jamanetworkopen.2019.18738.

Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes. Circulation. 2006; 113: 456-462. Available at: http://circ.ahajournals.org/content/113/3/456.full.pdf+html. Accessed January 14, 2020.

LaPar D J, Bhamidipati C M, et al. "Primary Payer Status Affects Mortality for Major Surgical Operations." Annals of Surgery. 2010; 252(3): 544-551.

LaPar D J, Stukenborg G J, et al "Primary Payer Status Is Associated With Mortality and Resource Utilization for Coronary Artery Bypass Grafting." Circulation. 2012; 126:132-139.

Lindenauer PK, Lagu T, Rothberg MB, et al. Income inequality and 30-day outcomes after acute myocardial infarction, heart failure, and pneumonia: retrospective cohort study. BMJ. 2013 Feb 14; 346:f521. doi: 10.1136/bmj.f521.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Pope GC, Ellis RP, Ash AS, et al. Diagnostic cost group hierarchical condition category models for Medicare risk adjustment. Final Report to the Health Care Financing Administration under Contract Number 500-95-048. 2000; http://www.cms.hhs.gov/Reports/downloads/pope_2000_2.pdf. Accessed February 25, 2020.

Pope GC, Kautter J, Ingber MJ, et al. Evaluation of the CMS-HCC Risk Adjustment Model: Final Report. 2011; https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/evaluation_risk_adj_model_2011.pdf. Accessed February 25, 2020.

Roshanghalb A, Mazzali C, Lettieri E. Multi-level models for heart failure patients' 30-day mortality and readmission rates: the relation between patient and hospital factors in administrative data. *BMC Health Serv Res*. 2019;19(1):1012. Published 2019 Dec 30. doi:10.1186/s12913-019-4818-2.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

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

- ☒ **Published literature**
- ☒ **Internal data analysis**
- ☐ **Other (please describe)**

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

The table below shows the final variables in the model in the testing dataset with associated odds ratios (OR) and 95 percent confidence intervals (CI).

Table 4. Adjusted Odds Ratios (ORs) and 95% Confidence Intervals (Cis) for the COPD Mortality Hierarchical Logistic Regression Model over Different Time Periods in the **Testing Dataset**

Variable	07/2016-06/2017 OR (95% CI)	07/2017-06/2018 OR (95% CI)	07/2018-06/2019 OR (95% CI)	07/2016-06/2019 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.04 (1.04-1.04)	1.04 (1.03-1.04)	1.03 (1.03-1.04)	1.04 (1.04-1.04)
History of mechanical ventilation	1.26 (1.21-1.32)	1.25 (1.19-1.31)	1.29 (1.23-1.36)	1.27 (1.23-1.30)
Metastatic cancer and acute leukemia (CC 8)	2.61 (2.45-2.77)	2.54 (2.37-2.71)	2.49 (2.31-2.67)	2.55 (2.45-2.65)
Lung and other severe cancers (CC 9)	1.70 (1.63-1.78)	1.60 (1.52-1.68)	1.59 (1.51-1.68)	1.64 (1.59-1.69)
Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other cancers and tumors; other respiratory and heart neoplasms (CC 10-13)	1.03 (0.99-1.07)	1.02 (0.98-1.07)	1.04 (0.99-1.09)	1.03 (1.01-1.06)

Variable	07/2016-06/2017 OR (95% CI)	07/2017-06/2018 OR (95% CI)	07/2018-06/2019 OR (95% CI)	07/2016-06/2019 OR (95% CI)
Other digestive and urinary neoplasms (CC 14)	0.86 (0.81-0.91)	0.84 (0.78-0.89)	0.88 (0.82-0.95)	0.86 (0.83-0.89)
Diabetes mellitus (DM) or DM complications (CC 17-19, 122-123)	0.97 (0.94-1.00)	0.94 (0.91-0.97)	0.98 (0.95-1.02)	0.96 (0.94-0.98)
Protein-calorie malnutrition (CC 21)	2.17 (2.10-2.25)	2.0 (2.00-2.16)	2.08 (1.99-2.16)	2.13 (2.09-2.18)
Morbid obesity; other endocrine/metabolic/nutritional disorders (CC 22, 25-26)	0.87 (0.84-0.91)	0.87 (0.84-0.91)	0.87 (0.83-0.92)	0.87 (0.85-0.89)
Other significant endocrine and metabolic disorders; disorders of fluid/electrolyte/acid-base balance (CC 23-24)	1.10 (1.07-1.14)	1.10 (1.06-1.14)	1.07 (1.03-1.12)	1.09 (1.07-1.12)
Other gastrointestinal disorders (CC 38)	0.86 (0.84-0.89)	0.86 (0.83-0.89)	0.8 (0.83-0.90)	0.86 (0.84-0.88)
Osteoarthritis of hip or knee (CC 42)	0.77 (0.74-0.81)	0.82 (0.78-0.86)	0.79 (0.75-0.83)	0.79 (0.77-0.82)
Other musculoskeletal and connective tissue disorders (CC 45)	0.88 (0.85-0.91)	0.84 (0.81-0.87)	0.84 (0.81-0.88)	0.86 (0.84-0.87)
Iron deficiency or other/unspecified anemias and blood disease (CC 49)	1.25 (1.21-1.29)	1.23 (1.19-1.28)	1.26 (1.22-1.31)	1.25 (1.23-1.28)
Dementia or other specified brain disorders (CC 51-53)	1.30 (1.25-1.34)	1.24 (1.19-1.28)	1.28 (1.23-1.34)	1.28 (1.25-1.30)
Drug/alcohol abuse, without dependence (CC 56)	0.90 (0.87-0.93)	0.82 (0.79-0.85)	0.86 (0.83-0.89)	0.86 (0.84-0.88)
Other psychiatric disorders (CC 63)	1.17 (1.13-1.21)	1.16 (1.12-1.20)	1.12 (1.08-1.16)	1.15 (1.13-1.17)
Hemiplegia, paraplegia, paralysis, functional disability (CC 70-74, 103-104, 189-190)	1.06 (1.00-1.13)	1.02 (0.96-1.09)	1.11 (1.03-1.19)	1.06 (1.02-1.10)
Polyneuropathy, mononeuropathy, and other neurological conditions/injuries (CC 81)	0.89 (0.86-0.92)	0.87 (0.84-0.90)	0.91 (0.87-0.95)	0.88 (0.86-0.90)
Respirator dependence/respiratory failure (CC 82-83)	0.93 (0.83-1.04)	0.89 (0.80-1.01)	0.92 (0.82-1.05)	0.92 (0.86-0.98)
Cardio-respiratory failure and shock (CC 84 plus ICD-10-CM codes R09.01 and R09.02, for discharges on or after October 1, 2015; CC 84 plus ICD-9-CM diagnosis codes 799.01 and 799.02, for discharges prior to October 1, 2015)	1.37 (1.33-1.42)	1.37 (1.32-1.42)	1.38 (1.33-1.44)	1.36 (1.34-1.39)
Congestive heart failure (CC 85)	1.23 (1.19-1.27)	1.24 (1.20-1.29)	1.21 (1.16-1.26)	1.23 (1.20-1.25)

Variable	07/2016-06/2017 OR (95% CI)	07/2017-06/2018 OR (95% CI)	07/2018-06/2019 OR (95% CI)	07/2016-06/2019 OR (95% CI)
Coronary atherosclerosis or angina (CC 88-89)	0.99 (0.96-1.02)	1.02 (0.99-1.06)	1.00 (0.96-1.04)	1.00 (0.98-1.02)
Hypertension and hypertensive disease (CC 94-95)	0.88 (0.85-0.92)	0.86 (0.83-0.90)	0.84 (0.80-0.87)	0.86 (0.84-0.88)
Specified arrhythmias and other heart rhythm disorders (CC 96-97)	1.08 (1.04-1.11)	1.08 (1.05-1.12)	1.14 (1.09-1.18)	1.09 (1.07-1.12)
Stroke (CC 99-100)	0.95 (0.89-1.01)	0.98 (0.92-1.04)	0.97 (0.90-1.04)	0.96 (0.93-1.00)
Vascular or circulatory disease (CC 106-109)	0.99 (0.96-1.02)	1.03 (1.00-1.07)	0.98 (0.94-1.02)	1.00 (0.98-1.02)
Fibrosis of lung or other chronic lung disorders (CC 112)	1.19 (1.14-1.23)	1.22 (1.17-1.27)	1.19 (1.14-1.25)	1.20 (1.17-1.23)
Asthma (CC 113)	0.71 (0.69-0.74)	0.70 (0.67-0.73)	0.65 (0.62-0.69)	0.71 (0.69-0.72)
Pneumonia (CC 114-116)	1.18 (1.14-1.22)	1.24 (1.20-1.29)	1.24 (1.19-1.29)	1.22 (1.20-1.25)
Pleural effusion/pneumothorax (CC 117)	1.24 (1.20-1.29)	1.31 (1.26-1.36)	1.24 (1.18-1.29)	1.26 (1.23-1.29)
Other respiratory disorders (CC 118)	0.74 (0.72-0.76)	0.74 (0.71-0.76)	0.75 (0.72-0.78)	0.74 (0.73-0.76)
Other retinal disorders (CC 125)	0.89 (0.84-0.93)	0.96 (0.91-1.00)	0.98 (0.93-1.03)	0.93 (0.90-0.96)
Other eye disorders (CC 128)	0.92 (0.89-0.95)	0.87 (0.84-0.91)	0.94 (0.91-0.98)	0.92 (0.90-0.94)
Other ear, nose, throat, and mouth disorders (CC 131)	0.80 (0.78-0.83)	0.78 (0.76-0.81)	0.81 (0.78-0.84)	0.80 (0.78-0.81)
Renal failure (CC 135-140)	1.15 (1.11-1.18)	1.14 (1.10-1.18)	1.16 (1.11-1.20)	1.15 (1.12-1.17)
Decubitus ulcer or chronic skin ulcer (CC 157-161)	1.39 (1.33-1.45)	1.33 (1.27-1.40)	1.38 (1.30-1.45)	1.37 (1.33-1.41)
Other dermatological disorders (CC 165)	0.90 (0.88-0.93)	0.91 (0.88-0.94)	0.92 (0.89-0.96)	0.91 (0.89-0.93)
Trauma (CC 166-168, 170-173)	1.05 (1.00-1.09)	1.04 (0.99-1.09)	1.08 (1.03-1.14)	1.05 (1.02-1.08)
Vertebral fractures without spinal cord injury (CC 169)	1.14 (1.08-1.21)	1.19 (1.11-1.26)	1.30 (1.22-1.39)	1.20 (1.15-1.24)
Major complications of medical care and trauma (CC 176-177)	0.90 (0.86-0.95)	0.91 (0.86-0.96)	0.92 (0.87-0.98)	0.91 (0.88-0.94)

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

Throughout this section, we present new SRF testing results based on the current testing dataset (2020); in addition, we show prior analyses included in the 2016 endorsement maintenance forms for comparison purposes.

Variation in prevalence of the factor across measured entities in 2020 and 2016 (Table 5)

SRFs	2020 Prevalence % (IQR)	2016 Prevalence % (IQR)
Dual	18.6% (11.3-28.0%)	19.1% (12.9-26.8%)
AHRQ Low SES	20.2% (7.40-37.1%)	15.2% (2.9-44.7%)

The prevalence of social risk factors in the COPD cohort varies widely across measured entities in 2020. The median percentage of dual eligible patients was 18.6% (IQR 11.3%-28.0%) and the median percentage of patients with an AHRQ SES index score adjusted for cost of living at the census block group level equal to or below 42.7 (lowest quartile) was 20.2% (IQR 7.40%-37.1%) in 2020. These results are consistent with the 2016 results presented above.

Comparison of observed mortality rates in patients with and without social risk in 2020 and 2016 (Table 6)

SRFs	2020 Observed Rate	2016 Observed Rate
Dual (vs. Non-Dual)	8.4% (vs. 8.4%)	7.2% (vs. 7.8%)
AHRQ Low SES (vs. SES score above 42.7)	8.1% (vs 8.6%)	7.1% (vs. 7.9%)

The patient-level observed COPD mortality rates are the same for dual-eligible patients (8.4%) compared with 8.4% for non-dual patients in 2020. Similarly, the mortality rate for patients with an AHRQ SES index score equal to or below 42.7 was 8.1% compared with 8.6% for patients with an AHRQ SES index score above 42.7 in 2020. Overall, the rates have increased in the COPD cohort since reported in 2016 which were based on data from July 2011- June 2014.

Incremental effect of SES variables in a multivariable model in 2020 and 2016

We examined the strength and significance of the SRF variables in the context of a multivariable model. When we include these variables in a multivariable model that includes all of the claims-based clinical variables, the effect size of each of these variables is small. In 2020, dual eligibility and the AHRQ SES index have effect sizes of 0.93 and 0.99 when added independently to the model. Furthermore, the effect size of each variable is attenuated (0.92 and 1.00 for dual eligible and SES) when both are added to the model.

We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model (Table 7).

Table 7

COPD Mortality Models	2020 C-Statistic
Base Model: risk-adjusted model using the original clinical risk variables selected for the 2020 CMS public report of the COPD mortality measure	0.73
Base Model plus AHRQ Low SES based on beneficiary residential 9-digit ZIP codes (SES9) as a social risk variable	0.73
Base Model plus dual eligibility (dual) as a social risk variable	0.73
Base Model plus SES9 and dual as social risk variables	0.73

Furthermore, we find that the addition of any of these variables into the hierarchical model has little to no effect on hospital performance. We examined the change in hospitals' RSMRs with the addition of any of these variables. The median absolute change in hospitals' RSMRs when adding a dual eligibility indicator is 0.009% (interquartile range [IQR] -0.011% – 0.007%) with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.999. The median absolute change in hospitals' RSMRs when adding a low AHRQ SES Index score indicator to the model is 0.080% (IQR -0.033% – 0.094%) with a correlation coefficient between RSMRs for each hospital with and without an indicator for a low AHRQ SES Index score adjusted for cost of living at the census block group level is 0.981.

Summary

Overall, we find that among the SRF variables that could be feasibly incorporated into this model, the relationship between dual-eligible status and AHRQ low SES is in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also find that the impact of any of these indicators is small to negligible on model performance and hospital-level results. Given the controversial nature of incorporating such variables into a risk-model, we do not support doing so in a case that is unlikely to affect hospital profiling. Given these empiric findings, ASPE's recommendation to not risk adjust publicly reported quality measures for SRFs, and complex pathways which could explain the relationship between SRFs and mortality (and do not all support risk-adjustment), CMS chose to not incorporate SRF variables in this measure.

References:

Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (ASPE). Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-based Purchasing Programs. 2020; https://aspe.hhs.gov/system/files/pdf/263676/Social-Risk-in-Medicare%E2%80%99s-VBP-2nd-Report.pdf. Accessed July 2, 2020.

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

Approach to assessing model performance

We computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the expanded cohort:

Discrimination Statistics

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.2)

Calibration Statistics

(3) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)

We tested the performance of the model for **the development dataset** described in section 1.7.

References:

Harrell FE and Shih YC, Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* **17** (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

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

Development and Validation Dataset:

1st half of randomly split development sample:

- c-statistic = 0.72
- Predictive ability (lowest decile %, highest decile %) = (1.52, 23.74)

2nd half of randomly split development sample:

- c-statistic = 0.723
- Predictive ability (lowest decile %, highest decile %) = (1.60, 23.78)

Results for the Testing Dataset

- C-statistic = 0.73
- Predictive ability (lowest decile %, highest decile %): (1.5, 23.9)

For comparison of model with and without inclusion of social risk factors, see above section.

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

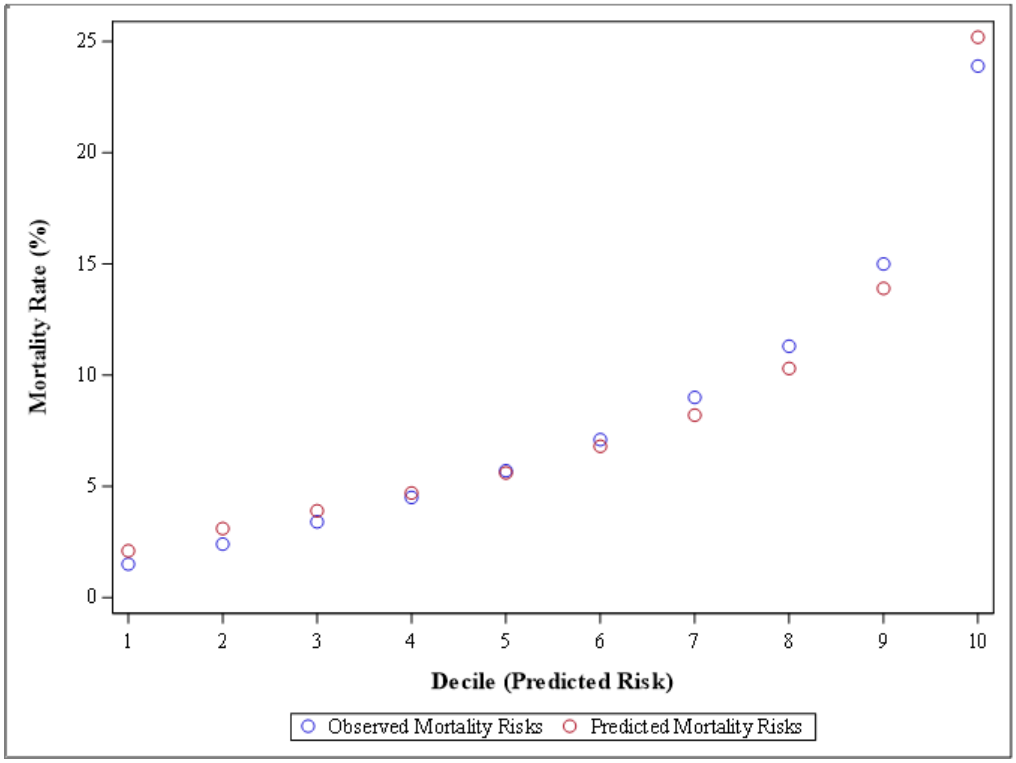
For the development cohort, the results are summarized below:

Development sample: Calibration: (-0.034, 0.985)

Validation sample: Calibration: (0.009, 1.004)

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

The risk decile plot (Figure 3) is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2016 – June 2019 (Testing Dataset).



2b3.9. Results of Risk Stratification Analysis:

N/A

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

Discrimination Statistics

The c-statistic of 0.73 indicate moderate model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates calibration of the model.

Risk Decile Plots

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates good discrimination of the model and good predictive ability.

Overall Interpretation

Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

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

-NA-

_____

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

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

The measure score is hospital-specific risk-standardized mortality rates. These rates are obtained as the ratio of predicted to expected mortality, multiplied by the national unadjusted rate. The "predicted" mortality (the numerator) is calculated using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of mortality. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are then transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" mortality (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are then transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimated the model coefficients using the years of data in that period.

We characterize the degree of variability by:

1) Reporting the distribution of RSMRs.
    a. For public reporting of the measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (because it is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 cases in the three-year period.
2) Providing the median odds ratio (MOR) (Merlo et al, 2006)
    a. The median odds ratio represents the median increase in the odds of mortality within 30 days of a COPD admission date on a single patient if the admission occurred at a higher risk hospital compared to a lower risk hospital. MOR quantifies the between hospital variance in terms of odds ratio, it is comparable to the fixed effects odds ratio.

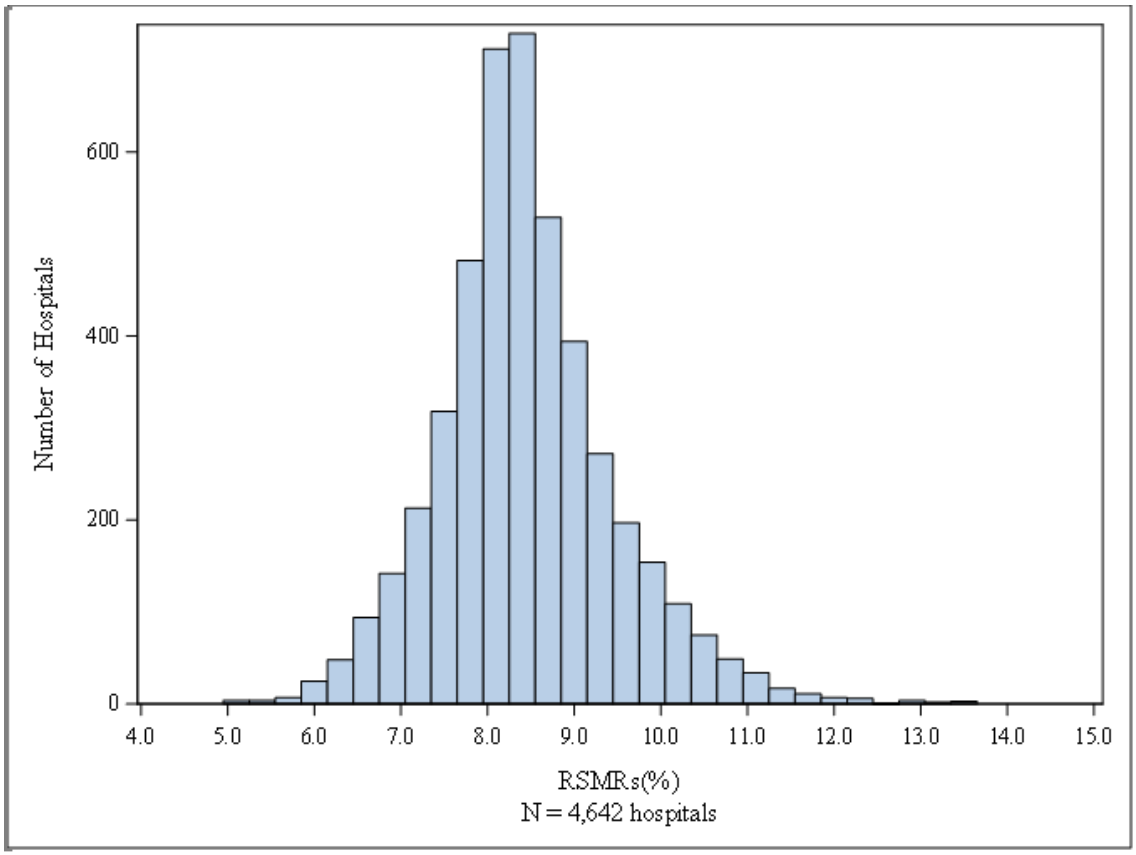
Reference

Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, Råstam L, Larsen K. (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health, 60(4):290-7.

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

Analyses of Medicare FFS data show substantial variation in RSMRs among hospitals.

Figure 4. Distribution (Histogram) Of Hospital-Level COPD RSMRs



Out of 4,642 hospitals in the measure cohort, 63 performed "better than the U.S. national rate," 3,554 performed "no different from the U.S. national rate," and 86 performed "worse than the U.S. national rate." 939 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

The median odds ratio was 1.26.

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

The median odds ratio suggests a meaningful increase in the risk of mortality if a patient is admitted with COPD at a higher risk hospital compared to a lower risk hospital. A value of 1.26 indicates that a patient's risk of mortality is 26% greater in a higher risk hospital than a lower risk hospital.

The variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for COPD. This evidence supports continued measurement to reduce the variation.

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

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

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

factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

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

N/A

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

N/A

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

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

The COPD mortality measure used claims-based data for development and testing. There was no missing data in the development and testing data.

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

N/A

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

N/A

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

N/A

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

Attachment:

3c. Data Collection Strategy

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

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

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

This measure uses administrative claims and enrollment data and as such, offers no data collection burden to hospitals or providers.

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

N/A

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting Hospital Compare https://www.medicare.gov/hospitalcompare/search.html? Hospital Compare https://www.medicare.gov/hospitalcompare/search.html? Payment Program Hospital Value Based Purchasing Program (HVBP) https://www.qualitynet.org/inpatient/hvbp

*cell intentionally left blank

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

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

Public Reporting
Program Name, Sponsor: Hospital Compare, Centers for Medicare and Medicaid Services (CMS)
Purpose: Under Hospital Compare and other CMS public reporting websites, CMS collects quality data from hospitals, with the goal of driving quality improvement through measurement and transparency by publicly displaying data to help consumers make more informed decisions about their health care. It is also intended to encourage hospitals and clinicians to improve the quality and cost of inpatient care provided to all patients. The data collected are available to consumers and providers on the Hospital Compare website at: https://www.medicare.gov/hospitalcompare/search.html. Data for selected measures are also used for paying a portion of hospitals based on the quality and efficiency of care, including the Hospital Value-Based Purchasing Program, Hospital-Acquired Condition Reduction Program, and Hospital Readmissions Reduction Program.
Payment Program
Program Name, Sponsor: Hospital Value-Based Purchasing (HVBP) Program, Centers for Medicare and Medicaid Services (CMS)
Purpose: The Hospital Value-Based Purchasing (VBP) Program is a CMS initiative that rewards acute-care hospitals with incentive payments for the quality of care they provide to people with Medicare. It was established by the Affordable Care Act of 2010 (ACA), which added Section 1886(o) to the Social Security Act. The law requires the Secretary of the Department of Health and Human Services (HHS) to establish a value-based purchasing program for inpatient hospitals. To improve quality, the ACA builds on earlier legislation—the 2003 Medicare Prescription Drug, Improvement, and Modernization Act and the 2005 Deficit Reduction Act. These earlier laws established a way for Medicare to pay hospitals for reporting on quality measures, a necessary step in the process of paying for quality rather than quantity.
Geographic area and number and percentage of accountable entities and patients included: More than 3,000 hospitals across the country are eligible to participate in Hospital VBP. The program applies to subsection (d)

hospitals located in the 50 states and the District of Columbia and acute-care hospitals in Maryland. More details about the Hospital VBP program are online at https://www.qualitynet.org/inpatient/hvbp.
The following hospitals are excluded from Hospital VBP:

- Hospitals and hospital units excluded from the Inpatient Prospective Payment System, such as psychiatric, rehabilitation, long-term care, children's, and cancer hospitals;
- Hospitals that are located in the state of Maryland participating in the Maryland All-Payer Model;
- Hospitals subject to payment reductions under the Hospital Inpatient Quality Reporting (IQR) Program;
- Hospitals cited by the Secretary of HHS for deficiencies during the performance period that pose an immediate jeopardy to patients' health or safety;
- Hospitals with an approved extraordinary circumstance exception specific to Hospital VBP; and
- Hospitals that do not meet the minimum number of cases, measures, or surveys required by Hospital VBP.

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

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

N/A. This measure is currently publicly reported.

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

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

The exact number of measured entities (acute care hospitals) varies with each new measurement period. For the period between 2016 – 2019, all non-federal short-term acute care hospitals (including Indian Health Service hospitals), critical access hospitals, and VA hospitals (4,642 hospitals) were included in the measure calculation. Only those hospitals with at least 25 COPD admissions were included in public reporting.

Each hospital receives their measure results in the Spring of each calendar year through CMS's QualityNet website. The results are then publicly reported on CMS's Hospital Compare website in July of each calendar year. Since the measure is risk standardized using data from all hospitals, hospitals cannot independently calculate their score.

However, CMS provides each hospital with several resources that aid in the interpretation of their results (described in detail below). These include Hospital-Specific Reports with details about every patient from their facility that was included in the measure calculation (for example, dates of admission and discharge, discharge diagnoses, outcome [died or not], transfer status, and facility transferred from). These reports facilitate quality improvement activities such as review of individual deaths and patterns of deaths; make visible to hospitals post-discharge outcomes that they may otherwise be unaware of; and allow hospitals to look for patterns that may inform quality improvement (QI) work (e.g. among patient transferred in from particular facilities). CMS also provides measure FAQs, webinars, and measure-specific question and answer inboxes for stakeholders to ask specific questions.

The Hospital-Specific Reports also provide hospitals with more detailed benchmarks with which to gauge their performance relative to peer hospitals and interpret their results, including comorbidity frequencies for their patients relative to other hospitals in their state and the country.

Additionally, the code used to process the claims data and calculate measure results is written in SAS (Cary, NC) and is provided each year to hospitals upon request.

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

During the Spring of each year, hospitals have access to the following list of updated resources related to the measure which is provided directly or posted publicly for hospitals to use:

1. Hospital-Specific Reports (HSR): available for hospitals to download from QualityNet in April/May of each calendar year; includes information on the index admissions included in the measure calculation for each facility, detailed measure results, and state and national results.
2. HSR User Guide: available with the HSR and posted on QualityNet; provides instructions for interpreting the results and descriptions of each data field in the HSR.
3. Mock HSR: posted on QualityNet; provides real national results and simulated state and hospital results for stakeholders who do not receive an HSR.
4. HSR Tutorial Video: A brief animated video to help hospitals navigate their HSR and interpret the information provided.
5. Public Reporting Preview and Preview Help Guide: available for hospitals to view from QualityNet in Spring of each calendar year; includes measure results that will be publicly reported on CMS's public reporting websites.
6. Annual Updates and Specification Reports: posted in April/May of each calendar year on QualityNet with detailed measure specifications, descriptions of changes made to the measure specifications with rationale and impact analysis (when appropriate), updated risk variable frequencies and coefficients for the national cohort, and updated national results for the new measurement period.
7. Frequently asked Questions (FAQs): includes general and measure-specific questions and responses, as well as infographics that explain complex components of the measure's methodology, and are posted in April/May of each calendar year on QualityNet.
8. The SAS code used to calculate the measure with documentation describing what data files are used and how the SAS code works. This code and documentation are updated each year and are released upon request beginning in July of each year.
9. Measure Fact Sheets: provides a brief overview of measures, measure updates, and are posted in April/May of each calendar year on QualityNet.

During the summer of each year, the publicly-reported measure results are posted on CMS's public reporting websites, a tool to find hospitals and compare their quality of care that CMS created in collaboration with organizations representing consumers, hospitals, doctors, employers, accrediting organizations, and other federal agencies. Measure results are updated in July of each calendar year.

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

Describe how feedback was obtained.

Question and Answer Inbox (Q&A)

The measured entities (acute care hospitals) and other stakeholders or interested parties submit questions or comments about the measure through an email inbox (CMSmortalitymeasures@yale.edu). Experts on measure specifications, calculation, or implementation, prepare responses to those inquiries and reply directly to the sender. We consider issues raised through the Q&A process about measure specifications or measure calculation in measure reevaluation.

Literature Reviews

In addition, we routinely scan the literature for scholarly articles describing research related to this measure. We summarize new information obtained through these reviews every 3 years as a part of comprehensive reevaluation as mandated by the Measure Management System (MMS) Blueprint.

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

Summary of Questions or Comments from Hospitals submitted through the Q & A process:

For the COPD mortality measure, we have received the following inquiries from hospitals since the last endorsement maintenance cycle:

1. Requests for detailed measure specifications including the ICD-9 and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
2. Requests for the SAS code used to calculate measure results;
3. Requests about the data source used to calculate the measure;
4. Questions about how transfers are handled in the measure calculation;
5. Requests for hospital-specific measure information such as HSRs; and
6. Requests for clarification of how inclusion and exclusion criteria are applied.

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

Summary of Question and Comments from Other Stakeholders:

For the COPD mortality measure, we have received the following feedback from other stakeholders since the last endorsement maintenance cycle:

3. Requests for detailed measure specifications including the narrative specifications for the measure, CC-to-ICD-9 code crosswalks, and ICD-9 and ICD-10 codes used to define the measure cohort or in the risk-adjustment model;
4. Requests for the data source and the SAS code used to calculate measure results;
5. Requests for clarification of how inclusion and exclusion criteria are applied;
6. Queries about how cohorts and outcomes are defined, including how planned readmissions are defined;
7. Questions about how transfers are handled in the measure calculation.

Summary of Relevant Publications from the Literature Review:

Since the last endorsement cycle, we have reviewed more than 350 articles related to mortality following COPD admissions. Relevant articles shared key themes related to: spillover effects of the COPD mortality measure on readmission rates for other conditions; considerations for additional risk adjustment variables, including social risk factors and other clinical comorbidities; association between public reporting of mortality rates and trends in mortality rates; potential unintended consequences of readmission measures on mortality outcomes; and, the clinical differences between different types of COPD.

Researchers have conducted considerable investigation of potential unintended consequences since the implementation of the Hospital Readmission Reductions Program. More specifically, the relationship between the implementation of readmission measures in the Hospital Readmissions Reduction Program (HRRP) and subsequent trends in their respective mortality rates has been studied.

Some studies have argued that since HRRP implementation, mortality for some conditions (including COPD) has increased, suggesting a potential unintended consequence that readmission measures may be incentivizing hospitals to not readily admit patients with COPD, and as a result, mortality rates increased (Samarghandi et al., 2019). However, empiric findings and other studies have found no apparent increase in COPD mortality (Ni et al., 2016; MedPAC, 2018; Stensland., 2019).

Given the importance of this potential issue on patient outcomes, CMS commissioned an independent group to investigate whether there have been increases in mortality rates after HRRP implementation. CMS found through this investigation that no sufficient evidence exists to suggest that mortality has increased because of

the HRRP readmission measures. CMS is committed to continuing to monitor trends in same-condition readmission and mortality rates through annual measure reevaluation and surveillance tasks.

References:

Medicare Payment Advisory Commission. Mandated report: The effects of the Hospital Readmissions Reduction Program. Washington, DC 07/18 2018.

Ni H, Xu J. COPD-related Mortality by Sex and Race Among Adults Aged 25 and Over: United States, 2000-2014. NCHS Data Brief. 2016(256):1-8.

Samarghandi A, Qayyum R. Effect of Hospital Readmission Reduction Program on Hospital Readmissions and Mortality Rates [published online ahead of print, 2019 Sep 18]. J Hosp Med. 2019;14:E25-E30. doi:10.12788/jhm.3302.

Stensland J. MedPAC evaluation of Medicare's Hospital Readmission Reduction Program: Update. In:2019.

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

Each year, issues raised through the Q&A or in the literature related to this measure are considered by measure and clinical experts. Any issues that warrant additional analytic work due to potential changes in the measure specifications are addressed as a part of annual measure reevaluation. If small changes are indicated after additional analytic work is complete, those changes are usually incorporated into the measure in the next measurement period. If the changes are substantial, CMS may propose the changes through rulemaking and adopt the changes only after CMS received public comment on the changes and finalizes those changes in the IPPS or other rule. There were no questions or issues raised by stakeholders requiring additional analysis or changes to the measure since the last endorsement maintenance cycle.

Improvement

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

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

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

The median hospital 30-day, all-cause, RSMR for the COPD mortality measure for the 3-year period between July 1, 2016 and June 30, 2019 was 8.3%. The median RSMR decreased by 0.7 absolute percentage points from July 2016-June 2017 (median RSMR: 8.6%) to July 2018-June 2019 (median: RSRR: 7.9%).

4b2. Unintended Consequences

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

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

We did not identify any unintended consequences during measure development, model testing, or re-specification. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

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

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0275 : Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 05)

0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization

0506 : Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization

1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

2888 : Accountable Care Organization Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions

3502 : Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure

3504 : Claims-Only Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Our measure cohort was heavily vetted by clinical experts, a technical expert panel, and a public comment period. Additionally, the measure, with the specified cohort, has been publicly reported since 2008. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

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

Appendix

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

Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Co.2 Point of Contact: Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-

Co.3 Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)

Co.4 Point of Contact: Doris, Peter, Doris.peter@yale.edu

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:

Laura M. Grosso, PhD, MPH

Peter Lindenauer, MD, MSc

Changqin Wang, MD, MS

Shantal Savage, BA

Jaymie Potteiger, MPH

Yun Wang, PhD

Zameer Abedin, BA

Lori L. Geary, MPH

Elizabeth E. Drye, MD, SM

Harlan M. Krumholz, MD, SM

Technical Expert Panel Members:

Darlene Bainbridge, RN, MS, NHA, CPHQ, CPHRM President/CEO, Darlene D. Bainbridge & Associates, Inc.

Robert A. Balk, MD, Director of Pulmonary and Critical Care Medicine, Rush University Medical Center

Dale Bratzler, DO, MPH, President and CEO, Oklahoma Foundation for Medical Quality

Scott Cerreta, RRT, Director of Education, COPD Foundation

Gerard J. Criner, MD, Director of Temple Lung Center and Divisions of Pulmonary and Critical Care Medicine, Temple University

Guy D'Andrea, MBA, President, Discern Consulting

Jonathan Fine, MD, Director of Pulmonary Fellowship, Research and Medical Education, Norwalk Hospital

David Hopkins, MS, PhD, Senior Advisor, Pacific Business Group on Health

Fred Martin Jacobs, MD, JD, FACP, FCCP, FCLM, Executive Vice President and Director, Saint Barnabas Quality Institute

Natalie Napolitano, MPH, RRT-NPS, Respiratory Therapist, Inova Fairfax Hospital

Russell Robbins, MD, MBA, Principal and Senior Clinical Consultant, Mercer

Working Group Panel Members:

David Au, MD, MS, Investigator, VA Puget Sound Healthcare System, Northwest HSR&D Center of Excellence; Associate Professor of Medicine, Department of Medicine, University of Washington

Jerry Krishnan, MD, PhD, Associate Professor, Departments of Medicine and Health Studies, University of Chicago; Director, Asthma Center and Refractory Obstructive Lung Disorders Clinic, University of Chicago

Todd Lee, PharmD, PhD, Associate Professor, Departments of Pharmacy Practice and Pharmacy Administration, University of Illinois at Chicago; Senior Investigator, Center for Management of Complex Chronic Care (CMC3), Hines VA Hospital

Richard Mularski, MD, MCR, MSHS, Clinical Investigator, Center for Health Research, Kaiser Permanente; Clinical Assistant Professor of Medicine, Oregon Health & Science University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2008

Ad.3 Month and Year of most recent revision: 07, 2019

Ad.4 What is your frequency for review/update of this measure? Annual

Ad.5 When is the next scheduled review/update for this measure? 2020

Ad.6 Copyright statement: N/A

Ad.7 Disclaimers: N/A

Ad.8 Additional Information/Comments: N/A