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 #: 3597

Corresponding Measures:

De.2. Measure Title: Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System

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

De.3. Brief Description of Measure: Risk-Standardized rate of acute, unplanned hospital admissions among Medicare Fee-for-Service (FFS) patients aged 65 years and older with multiple chronic conditions (MCCs).

1b.1. Developer Rationale: Hospital admission rates are an effective marker of ambulatory care quality. Hospital admissions from the outpatient setting reflect a deterioration in patients' clinical status and as such reflect an outcome that is meaningful to both patients and providers. Patients receiving optimal, coordinated high-quality care should use fewer inpatient services than patients receiving fragmented, low-quality care. Thus, high population rates of hospitalization may, at least to some extent, signal poor quality of care or inefficiency in health system performance.

Patients with MCCs are at high risk for hospital admission, often for potentially preventable causes, such as exacerbation of pulmonary disease. [1] Evidence from several Medicare demonstration projects suggests that care coordination results in decreased hospital admission rates among high-risk patients. [2] In addition, studies have shown that the types of ambulatory care clinicians this measure targets (for example, primary care providers and specialists caring for patients with MCCs) can influence admission rates through primary care clinician supply, continuity of care, and patient-centered medical home interventions such as team-based and patient-oriented care. [3-5] More recent studies speak directly to the positive effect that individual providers and group practices can have on lowering patients' hospital visit rates. In particular, they support that comprehensive and continuous care by individual providers can decrease care utilization. [6-7]

Thus, the anticipated net benefits of this unplanned hospital admission measure include, but are not limited to:

- Reduced numbers of hospitalizations and days hospitalized;
- Improved outpatient disease management;
- Reduced rates of complications, including mortality; and
- Cost savings resulting from fewer hospitalizations.

Overall, this measure will provide CMS with a valuable tool for assessing the performance of TINs (individual clinicians and groups of clinicians) in the MIPS program.

Citations

1. Abernathy K, Zhang J, Mauldin P, et al. Acute Care Utilization in Patients With Concurrent Mental Health and Complex Chronic Medical Conditions. J Prim Care Community Health. 2016;7(4):226-233.
2. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Aff (Millwood). 2012;31(6):1156-1166.
3. van Loenen T, van den Berg MJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Fam Pract. 2014;31(5):502-516.
4. Dale SB, Ghosh A, Peikes DN, et al. Two-Year Costs and Quality in the Comprehensive Primary Care Initiative. N Engl J Med. 2016;374(24):2345-2356.
5. Casalino LP, Pesko MF, Ryan AM, et al. Small primary care physician practices have low rates of preventable hospital admissions. Health Aff (Millwood). 2014;33(9):1680-1688
6. Bazemore, A., et al. (2018). "Higher Primary Care Physician Continuity is Associated With Lower Costs and Hospitalizations." Ann Fam Med. 16(6): 492-497.
7. O´Malley, A. S., et al. (2019). "New approaches to measuring the comprehensiveness of primary care physicians." Health Serv Res. 54(2): 356-366.

S.4. Numerator Statement: The outcome for this measure is the number of acute admissions per 100 person-years at risk for admission during the measurement period.

S.6. Denominator Statement: Patients included in the measure (target patient population)

The target patient population for the outcome includes Medicare FFS patients aged 65 years and older with multiple chronic conditions (MCCs).

Provider types included for measurement

- Primary care providers (PCPs): CMS designates PCPs as physicians who practice internal medicine, family medicine, general medicine, or geriatric medicine, and non-physician providers, including nurse practitioners, certified clinical nurse specialists, and physician assistants.
- Relevant specialists: Specialists covered by the measure are limited to those who provide overall coordination of care for patients with MCCs and who manage the chronic diseases that put the MCCs patients in the measure at risk of admission. These specialists were chosen with input from our Technical Expert Panel (TEP) and include cardiologists, pulmonologists, nephrologists, neurologists, endocrinologists, and hematologists/oncologists. However, as indicated below and in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.

Patient attribution

We begin by assigning each patient to the clinician most responsible for the patient's care. The patient can be assigned to a PCP, a relevant specialist, or can be left unassigned.

- A patient who is eligible for attribution can be assigned to a relevant specialist only if the specialist has been identified as "dominant". A specialist is considered "dominant" if they have two or more visits with the patient, as well as at least two more visits than any PCP or other relevant specialist. For example, if a patient saw a cardiologist four times in the measurement year, a PCP twice, and a nephrologist twice, the patient would be assigned to the cardiologist, having met the definition of "dominant" specialist. Note: Hematologists and oncologists are considered relevant specialists as they could be expected to manage MCCs patients' care, especially during periods of acute cancer treatment. However, as indicated below in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.
- There are two scenarios where a patient can be assigned to a PCP. First, the patient must have seen at least one PCP. The patient will then be assigned to the PCP with the highest number of visits as long as there is no relevant specialist who is considered "dominant." Second, if the patient has had more than one visit with a relevant specialist but no "dominant" specialist has been identified, and has two or more visits with a PCP, they will be assigned to that PCP.
- Finally, the patient will be unassigned if they only saw non-relevant specialists, if the patient has not seen a PCP and no "dominant" specialist can be identified, or if the patient has not had more than one visit with any individual PCP.

Patients are then assigned at the Taxpayer Identification Number (TIN) level, which includes solo clinicians and groups of clinicians who have chosen to report their quality under a common TIN.

- At the TIN level, patients are first assigned to the clinician (unique National Provider Identifier (NPI)/TIN combination since a given provider can be affiliated with more than one TIN) most responsible for their care (using the algorithm for individual clinician-level attribution above) and then patients "follow" their clinician to the TIN designated by the clinician. Patients unassigned at the individual clinician level continue to be unassigned at the TIN level.

(Note that an alternative attribution approach was considered and assessed as described in section 2b.3.11 of the testing attachment and in Appendix C of the attached methodology report.)

Person-time at risk

Persons are considered at risk for hospital admission if they are alive, enrolled in FFS Medicare, and not in the hospital during the measurement period. In addition to time spent in the hospital, we also exclude from at-risk time: 1) time spent in a SNF or acute rehabilitation facility; 2) the time within 10 days following discharge from a hospital, SNF, or acute rehabilitation facility; and 3) time after entering hospice care.

S.8. Denominator Exclusions: We exclude patients from the cohort for these reasons:

1. Patients without continuous enrollment in Medicare Part A or B during the measurement period.
2. Patients enrolled in hospice at any time during the year prior to the measurement year or at start of the measurement year.
3. Patients with no E&M visit to a MIPS eligible clinician.
4. Patients assigned to clinicians who do not participate in the QPP on the MIPS track.

5. Patients attributed to hematologists and oncologists.
6. Patients not at risk for hospitalization during the measurement year.

De.1. Measure Type: Outcome

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

S.20. Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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? Not applicable; this is not a paired or grouped measure.

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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 provides a logic model depicting that outpatient providers can decrease the rate of hospital admissions for patients with multiple chronic conditions (MCCs) by providing improved care coordination and continuity of care.
- The developer cited several studies that support the assertion that ambulatory care clinicians can influence admission rates through quality of care. Some examples listed in literature included supplementing patient telephone calls with in-person meetings; occasionally meeting in-person with providers; acting as a communication hub for providers; providing patients with evidence-based education; providing strong medication management; and providing comprehensive and timely transitional care after hospitalizations.

Question for the Committee:

- Is there at least one thing that the provider can do to achieve a change in the measure results?

Guidance from the Evidence Algorithm

Outcome Measure (Box 1) → Empirical evidence supports the relationship to at least one structure or process (Box 2) → 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.

- In the calendar year 2018, a total of 4,659,922 Medicare Fee-for-Service (FFS) MCC patients were attributed to 58,435 Merit-based Incentive Payment System (MIPS)-eligible tax identifier numbers (TINs).
- Overall, across all TINs with at least one attributed MCC patient, the risk-standardized acute admission rate (RSAAR) measure scores ranged from 17.5 to 131.5 per 100 person-years, with a median of 38.7 and an interquartile range of 36.5 to 41.8. The mean RSAAR and standard deviation were 39.5 ± 5.8 admissions per 100 person-years.

Disparities

- The final patient-level model includes two social risk factors: the Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index (lowest quartile vs. upper three quartiles) and an area-level measure of specialist physician density (lowest quartile vs. upper three quartiles).
    - The developer notes that in the multivariable model that included both of these social risk factors along with the demographic and clinical risk adjusters, they found relatively modest effects for the social risk factor variables. Their rate ratios and 95% confidence intervals were 1.08 (1.07, 1.08) for the AHRQ SES variable and 1.04 (1.04, 1.05) for the specialist density variable.
- The measure does not adjust for dual eligibility (DE).

Questions for the Committee:

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

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.

- MIPS program measure
- Supporting evidence
- No concerns - modeled after 2888

- I am not aware of any new studies/information that changes the evidence base for this measure
- Outcome measure to assess performance of clinical practice groups in Merit-based Incentive Payment System with at least 15 clinicians and 18 patients. This shows moderate to high ability to identify care quality
- Hospital admission rates for patients with multiple chronic conditions represent important contributions to well-being and a potential target for quality improvement

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?

- IQR 36.5-418 risk adjusted admissions per 100 person years. Substantial variation given risk adjustment
- Yes and disparities captured in SES index
- No concerns
- Yes. Variability presents an opportunity for improvement.
- Performance gap or opportunity for improvement demonstrated in the variation of risk-standardized acute admission rate (RSAAR) among 4044 clinician groups with at least one MCC patient. Disparities include two SRFs; AHRQ SES Index and an area-level measure of specialist physician density.
- There is a fairly broad range of admissions across the population with multiple chronic conditions, suggesting a potential area for quality improvement

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Scientific Methods Panel (SMP)

SMP Rating:

R: H-5; M-2; L-0; I-1

V: H-0; M-7; L-1; I-0

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.

Specifications:

- The measure is a modified version of an existing NQF-endorsed measure (NQF 2888), last reviewed for endorsement in 2016, which is being reviewed during the Fall 2020 cycle.

Reliability

- Reliability testing conducted at the measure score-level:
  - Signal-to-noise and intraclass correlation coefficient: Developers examined the distribution of mean and median reliabilities across patient volume for these clinician groups and reported a mean and median signal-to-noise reliability for the MIPS MCC measure of 0.453 and 0.451, respectively (range 0.038-0.999, Interquartile Range (IQR) 0.190-0.694). These results were for all MIPS TINs with at least one attributed MCC patient.
  - After applying a case minimum of 18 MCC patients per clinician group and the group size threshold of >15 clinicians per group, mean and median reliability for 4,044 TINs was 0.809 and 0.873, respectively (range 0.413-0.999, IQR 0.683-0.961)
  - The SMP did not raised any major concerns with reliability testing.

Validity

- Validity testing conducted at the measure score level:
  - New measure → only face validity was conducted.
  - The developer convened a TEP to provide input as to the conditions, groupings, and modeling. Public commenting was also requested. A survey of the TEP showed 83% of respondents agreed that the MIPS MCC admission measure can be used to distinguish good from poor quality of care.
    - Of 11 members assessing ability to distinguish good from poor, five of 11 (45%) somewhat agreed, five moderately agreed, and one strongly disagreed.
  - The final patient-level risk-adjustment model included 49 variables (47 demographic and clinical variables and two social risk factors). They used a negative binomial regression model with linear variance (NB-1) to risk adjust the measure. The model built off work done for the ACO MCC admission measure. Social risk factors included low AHRQ SES index and low physician-specialist density.

Questions for the Committee regarding reliability:

- *The Standing Committee should discuss the reliability results and confirm the case minimum of 18 MCC and group size threshold of >15 clinicians is appropriate.*
- *Does the Standing Committee 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 analyses for the measure. Does the Standing Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Does the Standing Committee 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 Standing 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?

- No issues
- Agree with SMP--no major concerns
- No concerns
- No concerns
- Signal-to-Noise used showed high reliability when clinical groups had at least 15 providers and minimum of 18 patients
- The definition of key variables used in the model for admission is less well-developed than other models in this iteration. If I am interpreting the model diagnostics appropriately, then only about 10% of the admission variation is accounted for by this model. This seems like a low number and is less than other prediction models in the other metrics in this readmission group.

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

- Reliability is inadequate for small groups/low patient volumes. Need info on minimum size group and patient load measure would be applied to.
- No concerns
- No concerns
- No
- No. High reliability if minimum threshold sizes used
- The model seems to provide adequate predictive ability and the observed to expected ratios are acceptable.

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

- No
- No concerns
- No concerns
- No
- No. Face validity assessed by technical expert panel (TEP) survey.
- The fact that the distribution of admission rates across multiple provider types has a similar median value of admission but very different distributions (Table 10) suggests some concerns about the validity of this metric.

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?

- Risk adjustment score is low 0.109 but within acceptable range
- Risk adjustment applied
- No concerns
- Yes
- The model included 49 risk-adjustment variables; 47 demographic/clinical variables and 2 social risk factors (Low AHRQ SES index and Low physician-specialist density)
- I have concerns about risk adjustment in this mix of heterogeneous providers and patients. Patients with multiple sets of risk factors for admission are not a homogeneous group, either among hospitals or among providers.

2b4-6. 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
- Minimal threats to validity
- No concerns
- No
- This measure should identify meaningful difference in quality.
- The fact that the distribution of admission rates varies fairly widely over a range of provider types makes it hard to compare this metric over hospitals and over provider-types. This suggests that hospitals with skewed ranges of provider types who care for these patients with multiple risk factors for admission.

Criterion 3. Feasibility

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

- The measure is coded by someone other than person obtaining original information.
- All data elements are in defined fields in a combination of electronic sources.
- There are no fees, licensing, or other requirements to use any aspect of this measure as specified.

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?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: 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?

- No concerns. Administrative data measure
- Feasible claims
- No concerns
- None
- No concerns. All data collected from defined fields in electronic sources already available
- Gathering the data to study this metric seems feasible, but the analysis may be very complex and possibly not meaningful

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.

Planned use in an accountability program? ☒ Yes ☐ No

Accountability program details:

- The measure is not currently publicly reported or used in an accountability application. CMS may propose this measure for use under the MIPS program.

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

- During development of the measure, Yale CORE and CMS provided performance results, data, and/or assistance with interpretation in several ways.
    - CORE recruited and met with a national TEP throughout measure development.
    - CORE hosted a public comment after reviewing the measure with the TEP.
    - CORE solicited public comments on the measure, and the developer took all comments into consideration, addressing them individually. Therefore, performance results and data were provided to members of the TEP and then made public through public comment.
    - CORE presented the measure to clinicians and practice managers in the voluntary Clinician Champions Program to elicit feedback.
    - CORE presented the measure at national conferences (e.g., CMS Quality Conference, AcademyHealth).

Additional Feedback:

- The measure was reviewed by NQF Measures Application Partnership (MAP) for the MIPS program and Medicare Shared Savings Program (MSSP) in the 2019-2020 cycle
    - The MAP conditionally supported pending NQF endorsement for MSSP.
    - The MAP expressed support for the measure's concept but did not support the measure for the MIPS program. The MAP proffered potential areas for mitigation.
        - Areas for mitigation:
            - The measure should apply to clinician groups, not to individual clinicians. This recommendation was partly driven by reliability results and partly by concerns that individual clinicians may lack the necessary resources and structural supports to effectively reduce the risk of admissions among their MCC patients compared with larger groups of clinicians.
            - The measure should use a higher reliability threshold (e.g., 0.7).
            - The measure developer should consider the NQF guidance on attribution and consider patient preference and selection as a method of attribution as that date becomes available.
            - The measure should undergo the NQF endorsement process.

    - The MAP suggested that rather than moving directly to this outcome measure, process measures that would get to the desired outcome might be an appropriate stepwise approach to increasing accountability.
    - The MAP Rural Health Workgroup noted that chronic conditions included in this measure are prevalent in rural residents. However, the Rural Health Workgroup does not believe this measure should be linked to payment for rural clinicians or clinician groups.
- The developer stated that in response to MAP feedback, the measure was defined for clinician groups and testing results were presented for clinician groups including at least 15 providers. Various cut-points for measure reliability were considered and tested, including a minimum reliability of 0.4 (at least 18 patients with MCC attributed per TIN). The reliability of 0.4 was selected to align with reliability for other MIPS measures and to optimize applicability of the measure to larger proportion of patients with MCCs, provider groups, and to optimize capture of the outcome. Patient attestation has not yet been tested.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- Measure has not been implemented and therefore does not have year-over-year performance data for review.

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 states the measure is not currently in use.

Potential harms

- The developer states the measure is not currently in use.

***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?

- New measure. No track record. Not clear from submission what will be given to providers.
- Potential use in the future
- No concerns
- Unsure
- This measure is not in use. Feedback obtained through YaleCORE and CMS. Also this measure reviewed by NQF MAP for the MIPS program and MSSP. The measure was adapted to clinical groups of at least 15 providers based on MAP feedback.
- My understanding is that this metric is rather new and has not been widely used or studied in a transparent manner. So much more feedback and critical appraisal is needed before accepting this measure as a valid metric. **Meaningful** and critical feedback

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 obvious harms.
- Not in use as yet
- No concerns
- Developer states the measure is not currently in use.
- The measure model demonstrates that it could be used to further the high-quality outcomes goals
- The usability of this metric is uncertain

Criterion 5: Related and Competing Measures

Related or competing measures

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

Harmonization

- The developer states that the measure specifications are harmonized to the extent possible. The only differences are for the CMS programs and measurement levels for which they are intended, e.g., the MIPS measure is attributed and scored for clinician groups under MIPS, and the ACO MCC admission measure is attributed and scored for Medicare ACOs.

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
- Harmonize as much as possible
- No concerns
- The developer states that the measure specifications are harmonized to the extent possible.
- Yes. Measure 2888 targeted for Medicare ACOs. Harmonization was done
- No apparent related or competing measures.

Public and Member Comments

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

- Comment by: American Medical Association

  The American Medical Association (AMA) appreciates the opportunity to comment on NQF Quality Positioning System (QPS) Measure #3597: Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment **System. While** this measure may be useful at the community or population level, the AMA believes it is not appropriate to attribute this utilization to an individual physician or practices. Our position is due to several factors. Specifically, the lack of evidence to support applying this measure to individual physicians or practices is particularly concerning. For example, the evidence form demonstrates that improved care coordination and programs focused on care management can lead to reductions in hospital admissions but requires multiple components such as a disease management program, health system, and/or hospital. We do not believe that sufficient evidence was provided to support the theory that physicians or practices, in the absence of some coordinated program or payment offset (e.g., care management fee), can implement structures or processes that can lead to improved outcomes for these patients. In addition, the measure developer did not provide a sufficient level of information to demonstrate how the attribution approach is linked to the evidence provided.

  We are also disappointed to see the minimum measure score reliability results of 0.413 for practices with at least 15 clinicians and 18 patients with multiple chronic conditions. We believe that measures must meet <u>minimum</u> acceptable thresholds of 0.7 for reliability.

Lastly, the AMA does not believe that the current risk adjustment model is adequate due to the deviance R-squared of 0.105 but appreciates that the measure developer included the Agency for Healthcare Research and Quality Socioeconomic Status Index and physician-specialist density as variables within the risk model.

The AMA requests that the Standing Committee carefully consider whether this measure meets the NQF measure evaluation criteria or if additional revisions are needed prior to endorsement.

- Comment by: Federation of American Hospitals

The Federation of American Hospitals (FAH) appreciates the opportunity to comment on Measure #3597, Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System. The FAH asks that the Standing Committee carefully consider whether the attribution methodology is reasonable and evidence-based.

The FAH is also concerned that even though the median reliability score was 0.873 for practices with at least 15 clinicians and 18 patients with multiple chronic conditions, reliability ranged from 0.413 to 0.999. 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).

In addition, the FAH appreciates that the developer included the Agency for Healthcare Research and Quality Socioeconomic Status Index and physician-specialist density as variables within the risk model. Unfortunately, the FAH remains concerned with the risk model's fit since the deviance R-squared was only 0.105. The FAH does not believe that the reasons for this result are adequately addressed and risk adjustment must be improved prior to re-endorsement.

As a result, the FAH requests that the Standing Committee carefully consider whether the measure as specified should be endorsed.

- Comment by: Anonymous
  - I strongly support this measure as well-coordinated outpatient care is key to admission prevention.
- Of the 1 NQF member who have submitted a support/non-support choice:
  - 0 support the measure
  - 1 does not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3597
Measure Title: Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System
Panel Member #5: NOTE: The title of this measure on the MIF form uses the wording "under the Merit-based Incentive Payment System". Is this measure only used for this Merit-based Incentive Payment System or is the intent to use the measure more widely than this system? If the latter is the case, then these words should be deleted from the measure title. If the former is the case, then the Denominator state should make explicit that only those Providers should be included in the measure calculation.
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:**

Panel Member #2: Medicare enrollment data and different other publicly-available data files including American Community Survey, Area Health Resource Files

Panel Member #6: Medicare enrollment data

Panel Member # 9: Medicare enrollment; durable medical equipment (DME) claims

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

Panel Member #8: Yes. The rationale for attribution and those specialists excluded is provided. Assumptions behind the decision are somewhat arbitrary (i.e. cancer specialists really are essentially PCP's for cancer patients and responsible for coordination of care) and the definition of no "dominant specialist" is clearly stated. but the definitions were made in alignment with TEP recommendations and is reasonable.

***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 #4: The cohort qualifying conditions include certain chronic conditions related to the nervous, endocrine, cardiovascular, respiratory, and digestive systems but exclude other chronic conditions related to the integumentary, skeletal, muscular, lymphatic, urinary, and reproductive systems. Why is that? If the rational is that the cohort qualifying conditions are those "that most increased risk of admission" then wouldn't a measure that includes the non-qualifying conditions be a useful comparison group for purposes of validity? In general the denominator exclusions are provider-centric rather than patient-centric which results in a denominator of 4.9 million out of a possible denominator of approximately 22 million (60% of 37 million FFS beneficiaries. The exclusions tend to have validity related rationales except the #4 Patients assigned to clinicians who do not participate in the QPP on the MIPS track exclusion which is only a program requirement. Is the measure only valid when used in a particular program?

Panel Member #8: None.

Panel Member #9: No concerns

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 Panel Member #5: Not Appliable--X

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

Submission document: Testing attachment, section 2a2.2

Panel Member #1 :S/N analysis

Panel Member #4: Developer used the formula for signal-to-noise reliability presented by Adams et al. and the formula for intraclass correlation coefficient (ICC) presented by Nakagawa et al. to calculate individual clinician-level and TIN-level reliability scores

Panel Member #6: Developers estimated the clinician-group-level reliability using signal-to-noise analysis. The variation between clinician-groups ('signal') comprises the total variation ('noise' and 'signal') in the outcome in this case because the reliability of any one clinician-group's measure score will vary depending on the number of patients. Clinician-groups with higher volume will tend to have more reliable scores, while those with lower volume will tend to have less reliable scores.

Panel Member #8: Signal -to -noise analysis was conducted to assess and intraclass correlation coefficient.

Panel Member #9: signal to noise using ICC was used to compare single providers to group practices

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

Submission document: Testing attachment, section 2a2.3

Panel Member #1: Examined reliability for All MIPS TINS with one attributed MCC patient: mean 0.453 median 0.451 IQR: 0.190-0.694 MIPS TINS with >1 clinician with group threshold of >15 clinicians: mean 0.580, median 0.648, IQR 0.238-0.926 MIPS TINS with >15 clinicians and minimum 18 MCC patients: mean 0.809 median 0.873, IQR 6.83-0.961. Reliability was good for TINS with >15 clinicians >_18 MCC patients. Marginal for TINS >15 clinicians without regard to number of patients. Not reliable for All MIPS TINS with one attributed clinician.

Panel Member #2: In the scenarios with case minimum of 18 patients and the group size threshold of >15 clinicians per group, mean and median reliability for 4,044 TINs was 0.809 and 0.873, respectively (range 0.413-0.999, IQR 0.683-0.961), which may be considered high reliability.

Panel Member #4: Using a minimum case volume of 18 and minimum group size of 15 the measure had high reliability.

Panel Member #7: Signal-to-Noise: The mean and median signal-to-noise reliability score was calculated for clinician-groups with at least 1 attributed patient. The mean reliability score was 0.453 and median reliability score was 0.451. Using a group size threshold of >15 clinicians per group, mean and median reliability for 6,326 TINs were 0.580 and 0.648, respectively. Adding a case minimum of 18 patients per clinician-group and the group size threshold of >15 clinicians per group, mean and median reliability for 4,044 TINs were much higher at 0.809 and 0.873.

Panel Member #8: The median signal-to-noise reliability was 0.453 for all MIPS TIN's with at least one attributed MCC patient (N=58,435) with an interquartile range of 0.190 to 0.694, calculated using one year of data. A split-half analysis is not provided. It should be noted that the range was 0.038 to 0.999, the mean was 0.451. When confined to those groups with greater than 15 clinicians (N=6,326), the mean and median reliability was 0.580 and 0.648 respectively with a range of 0.038 to 0.999 and an interquartile range of 0.238 to 0.926. And, when analysis was extended to those groups with greater than 15 providers and a case minimum of at least 18 patients, the median reliability is 0.873.

Panel Member #9: Results demonstrated that individual provider scores were not reliable but scores for group practices of 15 or more providers increased reliability significantly (0.7-0.9) Per the NQF MAP recommendations, this measure is only recommended for practices of 15 or more providers

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

Panel Member # 6: assuming use only for clinician-groups of 15 or more with 18 or more patients in the group.

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

Panel Member # 6: would score assuming use with clinician-groups with 15 or more clinicians

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

Panel Member # 6: would score if using clinician groups with at least 1 attributed patient

☒ **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: Insufficient because not clear what criteria for minimum number of clinicians or minimum number of patients/TIN will be required under standard of model of measure. If restricted to groups with 15 clinicians and 18 patients, I would rate reliability MODERATE.

Panel Member #2: See my explanation in #7.

Panel Member #4: This submission demonstrates integrity in the determination of case volumes and group size minimums for high reliability

Panel Member #5: Reliability testing was adequate; signal-to-noise ratios varied from 0.45 to 0.87 for different clinician groupings.

Panel Member #6: Results of signal-to-noise show high reliability when apply criteria of at least 15 clinicians and 18 patients. Would like further explanation of how "signal" comprises total variation in this measure (no "noise")?

Panel Member #7: Among all MIPS TINs with at least one attributed MCC patient (n=58,435 TINs), mean and median signal-to-noise reliability for the MIPS MCC measure was 0.453 and 0.451, respectively (range 0.038-0.999, Interquartile Range (IQR) 0.190-0.694) (Table 1). Since the measure is intended to be used for clinician groups, we also calculated reliability for the measure when

applied to TINs with more than one clinician. To define practice group size, we calculated a count of NPIs associated with each TIN. The NPI count per TIN was based on any NPI associated with that TIN, irrespective of NPI specialty or MIPS eligibility. Using a group size threshold of >15 clinicians per group, mean and median reliability for 6,326 TINs was 0.580 and 0.648, respectively (range 0.038-0.999, IQR 0.238-0.926) (Table 1). We then examined the distribution of mean and median reliabilities across patient volume for these clinician groups and identified a case minimum of 18 MCC patients per clinician group as providing adequate reliability. Using this case minimum and the group size threshold of >15 clinicians per group, mean and median reliability for 4,044 TINs was 0.809 and 0.873, respectively (range 0.413-0.999, IQR 0.683-0.961) (Table 1).

Panel Member #8: The reliability meets criteria for high IF confined to the subset described above, namely those groups with more than 15 providers and with at least 18 patients with MCCs. For other populations, the reliability of the measure score is low

Panel Member #9: High when used for group practices of at least 15 providers

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1: NONE

Panel Member #2: None.

Panel Member #4: I think the measure should be specified independent of its use in a particular program, which should be more of an implementation consideration unless program participation is essential to measure validity.

Panel Member #5: List of exclusions seemed rather long and may substantially reduce the number of Taxpayer Identification Number (TIN) groups/practices included in the analyses and reporting. Table associated with item 2b2.2 seems to indicate that about 40% of beneficiaries are eliminated from the analyses due to these exclusion.

Panel Member #7: None

Panel Member #8: Approximately 2.6 million of the original 7.7.975 million patients were excluded due to exclusion criteria. Most common ones were the attribution to hematologists and oncologists, assignment to clinicians that did not participate in the QPP on the MIPS track, and those without continuous enrollment during the measurement period. Valid explanations are provided for the exclusions, consistent with the eligibility criteria mentioned above.

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: Data was from 2013-2015 were used for measure development and testing, with 2018 data used for ICD-10 testing.

Panel Member #2: I think the measure as computed should have reasonable ability to identify meaningful difference. But I am not very clear about how exactly the median rate ratio or MRR was calculated. As outlined in 2b4.1, it was calculated by taking all possible combinations of providers, always comparing the higher-risk provider to the lower-risk provider and the MRR is interpreted as a traditional rate ratio would be. This will be a large number of combinations of pairs of providers if we don't have any specific thresholds for higher- and lower-risk, and simply forms pairs with higher risk and lower risk providers.

Panel Member #4: None

Panel Member #6: The risk-standardized measure scores had a median value of 40.4 and mean value of 38.9 (standard deviation +/-4.2) admissions per 100 person-years. The percentiles of the distribution ranged from 20.4 (min) to 98.7 (max) and 26.7 (1st) to 65.6 (99th). Across the 4,044 clinician groups who had at least one MCC patient (not sure if this is right as the data presented

indicates 18 patients?), RSAAR measure scores, including adjustment for the social risk factors of AHRQ SES Index, and physician-specialist density, ranged from 20.4 to 98.7 per 100 person-years, with a median of 40.4 and an IQR of 36.0 to 45.2. **The developers I believe INCORRECTLY STATE that this indicates that "after adjustment half of Medicare patients with multiple chronic conditions had between 36 and 45 acute care visits in a year."** *I don't think that is quite possible, I believe they are referring to per 100 person-years.* The MRR value of 1.27 indicates that a patient has a 27% higher admission rate if the patient was attributed to a higher-risk clinician-group compared with a lower-risk clinician-group indicating that the impact of quality on the outcome rate is meaningful. Overall, the results suggest that there is substantial room to reduce the number of admissions.

Panel Member #8: The risk-standardized measure scores had a median value of 40.4 admissions per 100-person years, with a minimum of 20.4 and a maximum of 98.7. 1st percentile was 26.7%, 10th percentile 32.6, 90th percentile 50.2. Thus there is meaningful differences in performance.

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 #4: None

Panel Member #8: Not applicable.

Panel Member #6: N/A

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: NONE

Panel Member #2: No concern.

Panel Member #8: 0.44% of patients had missing data for one or both social risk factors, usually due to zip codes that could not provide data. This did not affect the results.

16. **Risk Adjustment**

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

Panel Member #6: 49 risk factors

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

Panel Member #1: AHRQ SES Index and physician-specialty density.

Panel Member #5: ZIP code level—Area Deprivation Index (ADI) from Census data (2009-2013)

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 N/A

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

Panel Member #5: See previous comments

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

Panel Member #1: Basic variation of standard CMS risk model, with variables for 9 conditions that trigger inclusion in the category, HCC comorbidities and other variable. Model explains 10.8% of variance, low but consistent with other risk adjustment models that have been approved. Risk deciles show a well calibrated model, although the presentation could have been clearer. While inclusion of SRF do not change results substantially from multivariate model adjusted for demographic and clinical variables, decision was made to include two SRF variables.

Panel Member #2: Explained in great detail, the developer adopted a negative binomial regression with linear variance (NB-1) model, which tested for demographic, clinical and SES/SRF variables. I am satisfied with all of the results except that I am not sure if the deviance R-squared of approximately 0.10 across different versions of the model with demographic and clinical risk factors in the Development and Validation subsamples and the 2015 Medicare MCC Full Sample (both with and without adding the AHRQ SES Index and physician-specialist density variables to the model) is the appropriate one.

Panel Member #4: Social risk factors are well conceptualized.

Panel Member #5: Risk adjustment was generally adequate, though not exceptional.

Panel Member #6: The final patient-level risk-adjustment model included 49 variables (47 demographic and clinical variables and 2 social risk factors). They used a negative binomial regression model with linear variance (NB-1) to risk adjust the measure. The model built off work done for the ACO MCC admission measure. Social risk factors included low AHRQ SES index and low physician-specialist density. It is interesting to note that the rationale for inclusion is somewhat in conflict for the rationale NOT to include social risk factors in all of the readmission measures from same developer. The developers indicate that CMS included these variables in the MIPS MCC measure because clinicians working in the community have a limited ability to influence these community-based contextual factors that affect admission risk. The clinician-group MCC conceptual model acknowledges that low SES influences admission risk. This is in direct conflict to the set of 30-day condition specific readmissions measures submitted by this developer measured at the hospital level. Is it true that hospitals have ability to influence community-based contextual factors but Clinician Groups do not? The developers also outline unintended consequences for not adjusting, including 1) could translate into downward Medicare payment adjustments for providers serving patients with social risk factors; 2) not adjusting might further reduce resources among providers already facing largest constraints; 3) providers may have an incentive to reduce access to care for vulnerable patients if they anticipate a poor score. Again, it is not clear how this logic applies to clinician groups but NOT to hospitals as in the readmissions measures. I would also note that in the multivariate model, dual eligible status was strongly predictive of admissions in the bivariate model (RR-1.42) and had an even stronger relationship to admissions (1.12) in the multivariate model including other clinical risk factors than the low SES (1.07) and low specialist density (1.04) but was NOT included in the final model. The model was evaluated using deviance R-squared, which was 0.105 indicating the model explains 10.5% of variation in admission rates. This is NOT indication of a very strong model.

Panel Member #8: The risk adjustment model utilized 49 variables, demographic (age), 46 clinical (diagnosis groupers and functional status), and 2 social risk variables (SES index and specialist density). A deviance R-squared was utilized for assessing the model performance. Predicted to measured number of admissions performed well across four quartiles.

***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: Face validity via feedback and comments received from TEP and public Association with selected measures of ACO performance, notably risk standardized all condition readmission (corr=0.42) and diabetes hemoglobin poor control (corr 0.18). No substantial correlation found with timeliness of appointments, care and information, access to specialists or controlling high blood pressure.

Panel Member #4: Face validity testing should be both transparent and systematic. I do not think the process used meets those criteria.

Panel Member #6: Validity was assessed by external stakeholders and experts, as well as a technical expert panel (TEP). TEP members were asked to assess the face validity of the final measure specification by confidentially reporting their agreement with following two statements. "The risk-standardized acute admission rates obtained from the MIPS MCC admission measure as specified: 1) Can be used to distinguish good from poor quality of care provided to MCC patients by TINs reporting under MIPS." And 2) Will provide TINs reporting under MIPS with information that can be used to improve their quality of care for MCC patients."

Panel Member #8: For face validity, a TEP was convened an provided expert panel input as to the conditions, groupings, modeling. Public commenting was also requested. A survey of the TEP showed 83% of respondents agreed that the MIPS MCC admission measure can be used to distinguish good from poor quality of care.

Panel Member #9: TEP input and Likert scale survey in addition to public comments

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: Face validity via feedback and comments received from TEP and public appears moderate, with some disagreement about use. Of 11 member assessing ability to distinguish good from poor, 5 of 11 somewhat agreed, 5 moderately agreed, 1 strongly disagreed.

Panel Member #2: Measure validity is demonstrated through assessment of face validity through (i) a technical expert panel (TEP) and use of established measure development guidelines and (ii) establishing validity as indicated by established measure development guidelines (e.g., NQF guidance for outcome measures, CMS MMS guidance, and guidance articulated in the American Heart Association scientific statement entitled, "Standards for Statistical Models Used for Public Reporting of Health Outcomes.).

Panel Member #4: The degree of consensus was moderate to low.

Panel Member #6: Of 17 TEP members, 12 responded. On question 1, 2 strongly or moderately disagreed and 10 somewhat or moderately agreed (83%). On question 2, 3 moderately or somewhat disagreed (25%), 3 somewhat agreed (25%) and and 6 moderately or strongly agreed (50%). Overall, these results support moderate measure score validity.

Panel Member #8: No empirical testing performed. Face validity formed the basis of the evaluation.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

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

☒ **Yes**

☒ **No**

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

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

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

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

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

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

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

Panel Member #1: Face validity is sufficiently established for initial endorsement.

Panel Member #2: I think the measure developer has explained well the face validity of the proposed measure, and I have no concern.

Panel Member #4: The rating of low is based on both the method and the result.

Panel Member #5: Developer demonstrated an effort to risk adjust the measure including the use of socio-demographic variables to create valid measure score.

Panel Member #6: The validity test results show moderate face validity of the model.

Panel Member #7: Rationale for variables included variation in measure score. TEP only review of measure score. TEP only review of measure score

Panel Member #8: Face validity alone provides the sole test of validity.

Panel Member #9: I struggled to choose between low or moderate. While I believe the measure developers performed validation within the confines of accepted processes, I would have liked to see more convincing evidence that this actually measures quality of care. Possibly more literature supporting increased admissions demonstrates poor quality.

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 #6: See comments related to evaluating the risk adjustment findings related to social risk factors.

Panel Member #8: This measure has low reliability across all groups and the importance of size of practice and number of patients is outlined above. Measure should be applied only to those populations. Validity is totally face validity from expert TEP without empirical testing.

Panel Member #9: Given the results in the submission I would only more forward on recommending this as a "trial-use" measure and for practices with 15 or more providers before full endorsement. I'd like to see more convincing evidence that this actually measures quality of care.

Developer Submission

NQF #: 3597

Corresponding Measures:

De.2. Measure Title: Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System

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

De.3. Brief Description of Measure: Risk-Standardized rate of acute, unplanned hospital admissions among Medicare Fee-for-Service (FFS) patients aged 65 years and older with multiple chronic conditions (MCCs).

1b.1. Developer Rationale: Hospital admission rates are an effective marker of ambulatory care quality. Hospital admissions from the outpatient setting reflect a deterioration in patients' clinical status and as such reflect an outcome that is meaningful to both patients and providers. Patients receiving optimal, coordinated high-quality care should use fewer inpatient services than patients receiving fragmented, low-quality care. Thus, high population rates of hospitalization may, at least to some extent, signal poor quality of care or inefficiency in health system performance.

Patients with MCCs are at high risk for hospital admission, often for potentially preventable causes, such as exacerbation of pulmonary disease. [1] Evidence from several Medicare demonstration projects suggests that care coordination results in decreased hospital admission rates among high-risk patients. [2] In addition, studies have shown that the types of ambulatory care clinicians this measure targets (for example, primary care providers and specialists caring for patients with MCCs) can influence admission rates through primary care clinician supply, continuity of care, and patient-centered medical home interventions such as team-based and patient-oriented care. [3-5] More recent studies speak directly to the positive effect that individual providers and group practices can have on lowering patients' hospital visit rates. In particular, they support that comprehensive and continuous care by individual providers can decrease care utilization. [6-7]

Thus, the anticipated net benefits of this unplanned hospital admission measure include, but are not limited to:

- Reduced numbers of hospitalizations and days hospitalized;
- Improved outpatient disease management;
- Reduced rates of complications, including mortality; and
- Cost savings resulting from fewer hospitalizations.

Overall, this measure will provide CMS with a valuable tool for assessing the performance of TINs (individual clinicians and groups of clinicians) in the MIPS program.

Citations

1. Abernathy K, Zhang J, Mauldin P, et al. Acute Care Utilization in Patients With Concurrent Mental Health and Complex Chronic Medical Conditions. J Prim Care Community Health. 2016;7(4):226-233.
2. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Aff (Millwood). 2012;31(6):1156-1166.

3. van Loenen T, van den Berg MJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Fam Pract. 2014;31(5):502-516.
4. Dale SB, Ghosh A, Peikes DN, et al. Two-Year Costs and Quality in the Comprehensive Primary Care Initiative. N Engl J Med. 2016;374(24):2345-2356.
5. Casalino LP, Pesko MF, Ryan AM, et al. Small primary care physician practices have low rates of preventable hospital admissions. Health Aff (Millwood). 2014;33(9):1680-1688
6. Bazemore, A., et al. (2018). "Higher Primary Care Physician Continuity is Associated With Lower Costs and Hospitalizations." Ann Fam Med. 16(6): 492-497.
7. O´Malley, A. S., et al. (2019). "New approaches to measuring the comprehensiveness of primary care physicians." Health Serv Res. 54(2): 356-366.

S.4. Numerator Statement: The outcome for this measure is the number of acute admissions per 100 person-years at risk for admission during the measurement period.

S.6. Denominator Statement: Patients included in the measure (target patient population)

The target patient population for the outcome includes Medicare FFS patients aged 65 years and older with multiple chronic conditions (MCCs).

Provider types included for measurement

- Primary care providers (PCPs): CMS designates PCPs as physicians who practice internal medicine, family medicine, general medicine, or geriatric medicine, and non-physician providers, including nurse practitioners, certified clinical nurse specialists, and physician assistants.
- Relevant specialists: Specialists covered by the measure are limited to those who provide overall coordination of care for patients with MCCs and who manage the chronic diseases that put the MCCs patients in the measure at risk of admission. These specialists were chosen with input from our Technical Expert Panel (TEP) and include cardiologists, pulmonologists, nephrologists, neurologists, endocrinologists, and hematologists/oncologists. However, as indicated below and in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.

Patient attribution

We begin by assigning each patient to the clinician most responsible for the patient's care. The patient can be assigned to a PCP, a relevant specialist, or can be left unassigned.

- A patient who is eligible for attribution can be assigned to a relevant specialist only if the specialist has been identified as "dominant". A specialist is considered "dominant" if they have two or more visits with the patient, as well as at least two more visits than any PCP or other relevant specialist. For example, if a patient saw a cardiologist four times in the measurement year, a PCP twice, and a nephrologist twice, the patient would be assigned to the cardiologist, having met the definition of "dominant" specialist. Note: Hematologists and oncologists are considered relevant specialists as they could be expected to manage MCCs patients' care, especially during periods of acute cancer treatment. However, as indicated below in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.
- There are two scenarios where a patient can be assigned to a PCP. First, the patient must have seen at least one PCP. The patient will then be assigned to the PCP with the highest number of visits as

long as there is no relevant specialist who is considered "dominant." Second, if the patient has had more than one visit with a relevant specialist but no "dominant" specialist has been identified, and has two or more visits with a PCP, they will be assigned to that PCP.

- Finally, the patient will be unassigned if they only saw non-relevant specialists, if the patient has not seen a PCP and no "dominant" specialist can be identified, or if the patient has not had more than one visit with any individual PCP.

Patients are then assigned at the Taxpayer Identification Number (TIN) level, which includes solo clinicians and groups of clinicians who have chosen to report their quality under a common TIN.

- At the TIN level, patients are first assigned to the clinician (unique National Provider Identifier (NPI)/TIN combination since a given provider can be affiliated with more than one TIN) most responsible for their care (using the algorithm for individual clinician-level attribution above) and then patients "follow" their clinician to the TIN designated by the clinician. Patients unassigned at the individual clinician level continue to be unassigned at the TIN level.

(Note that an alternative attribution approach was considered and assessed as described in section 2b.3.11 of the testing attachment and in Appendix C of the attached methodology report.)

Person-time at risk

Persons are considered at risk for hospital admission if they are alive, enrolled in FFS Medicare, and not in the hospital during the measurement period. In addition to time spent in the hospital, we also exclude from at-risk time: 1) time spent in a SNF or acute rehabilitation facility; 2) the time within 10 days following discharge from a hospital, SNF, or acute rehabilitation facility; and 3) time after entering hospice care.

S.8. Denominator Exclusions: We exclude patients from the cohort for these reasons:

1. Patients without continuous enrollment in Medicare Part A or B during the measurement period.
2. Patients enrolled in hospice at any time during the year prior to the measurement year or at start of the measurement year.
3. Patients with no E&M visit to a MIPS eligible clinician.
4. Patients assigned to clinicians who do not participate in the QPP on the MIPS track.
5. Patients attributed to hematologists and oncologists.
6. Patients not at risk for hospitalization during the measurement year.

De.1. Measure Type: Outcome

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

S.20. Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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? Not applicable; this is not a paired or grouped measure.

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_MIPS_MCC_EvidenceAttachment_for_CMS_Review-637418977375542223.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.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Not previously endorsed

Measure Title: Clinician Group Risk-Standardized Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Not a composite measure

Date of Submission: To be updated upon submission to NQF (anticipated November 2020 per NQF project timeline)

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

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

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Outcome:

☐ Patient-reported outcome (PRO):

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

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

☐ Process:

☐ 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.

Hospital admission rates are an effective marker of ambulatory care quality. Hospital admissions from the outpatient setting reflect a deterioration in patients' clinical status and as such reflect an outcome that is meaningful to both patients and providers. Patients receiving optimal, coordinated high-quality care should use fewer inpatient services than patients receiving fragmented, low-quality care. Thus, high population rates of hospitalization may signal poor quality of care or inefficiency in health system performance. Furthermore, these effects may be exacerbated in disadvantaged areas. [1]

Patients with MCCs are at high risk for hospital admission, often for potentially preventable causes, such as exacerbation of pulmonary disease. [2] Evidence from several Medicare demonstration projects suggests that care coordination results in decreased hospital admission rates among high-risk patients. [3] In addition, studies have shown that the types of ambulatory care clinicians this measure targets (for example, primary care providers and specialists caring for patients with MCCs) can influence admission rates through primary care clinician supply, continuity of care, medication prescribing and dispensing interventions, as well as patient-centered medical home interventions such as team-based care, home visits, and patient-oriented care. [4-11] Other studies speak directly to the positive effect that individual providers and group practices can have on lowering patients' hospital visit rates. In particular, they support that comprehensive and continuous care by individual providers can decrease care utilization. [12-13]

Given evidence that ambulatory care clinicians can influence hospital admission rates through optimal care and coordination, this measure will incentivize quality improvement efforts leading to improved patient outcomes.

Citations:

Jencks, S. F., et al. (2019). "Safety-Net Hospitals, Neighborhood Disadvantage, and Readmissions Under Maryland's All-Payer Program: An Observational Study." Ann Intern Med. doi: 10.7326/M16-2671

Abernathy K, Zhang J, Mauldin P, et al. Acute Care Utilization in Patients With Concurrent Mental Health and Complex Chronic Medical Conditions. J Prim Care Community Health. 2016;7(4):226-233.

Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Aff (Millwood). 2012;31(6):1156-1166.

van Loenen T, van den Berg MJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Fam Pract. 2014;31(5):502-516.

Dale SB, Ghosh A, Peikes DN, et al. Two-Year Costs and Quality in the Comprehensive Primary Care Initiative. N Engl J Med. 2016;374(24):2345-2356.

Casalino LP, Pesko MF, Ryan AM, et al. Small primary care physician practices have low rates of preventable hospital admissions. Health Aff (Millwood). 2014;33(9):1680-1688

Matzke GR, Moczygemba LR, Williams KJ, Czar MJ, Lee WT. Impact of a pharmacist–physician collaborative care model on patient outcomes and health services utilization. American Journal of Health-System Pharmacy. 2018;75(14):1039-1047

Ruiz S, Snyder LP, Rotondo C, Cross-Barnet C, Colligan EM, Giuriceo K. Innovative Home Visit Models Associated With Reductions In Costs, Hospitalizations, And Emergency Department Use. Health Affairs. 2017;36(3):425-432

Edwards ST, Saha S, Prentice JC, Pizer SD. Preventing Hospitalization with Veterans Affairs Home-Based Primary Care: Which Individuals Benefit Most? Journal of the American Geriatrics Society. 2017;65(8):1676-1683

Krumme AA, Glynn RJ, Schneeweiss S, et al. Medication Synchronization Programs Improve Adherence To Cardiovascular Medications And Health Care Use. Health Aff (Millwood). 2018;37(1):125-133. doi:10.1377/hlthaff.2017

Gabriel M, Powers C, Encinosa W, Bynum J. E-Prescribing and Adverse Drug Events: An Observational Study of the Medicare Part D Population With Diabetes. Medical care. 2017;55

Bazemore, A., et al. (2018). "Higher Primary Care Physician Continuity is Associated With Lower Costs and Hospitalizations." Ann Fam Med. 16(6): 492-497.

O'Malley, A. S., et al. (2019). "New approaches to measuring the comprehensiveness of primary care physicians." Health Serv Res. 54(2): 356-366.

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.)

****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.

As noted above, hospital admission rates are an effective marker of ambulatory care quality. Patients receiving optimal, coordinated high-quality care should use fewer inpatient services than patients receiving fragmented, low-quality care. Thus, high population rates of hospitalization may signal poor quality of care or inefficiency in health system performance.

There is strong evidence supporting the assertion that ambulatory care clinicians can influence admission rates through quality of care. [1-7] For example, Brown et al. (2012) pointed to four Medicare Coordinated Care Demonstration programs that reduced hospitalizations for high-risk patients by 13-30 per 100 beneficiaries per year (8-33% of hospitalizations). Brown et al. highlighted six program features that were associated with successfully reducing hospitalizations: 1) supplementing patient telephone calls with in-person meetings; 2) occasionally meeting in person with providers; 3) acting as a communication hub for providers; 4) providing patients with evidence-based education; 5) providing strong medication management; and 6) providing comprehensive and timely transitional care after hospitalizations. [8] In addition, van Loenen et al. found that higher levels of provider continuity decreased the risk of avoidable hospitalizations for ambulatory care-sensitive conditions (ACSCs) and chronic diseases, regardless of country and age group. [9] Favorable results were also shown by Dorr et al. (2000), Levine et al. (2012), Littleford et al. (2010), and Zhang et al. (2008). [10-13] Additionally, using

Medicare data, Bazemore et al. (2018) found that primary care physicians with higher patient continuity-of-care scores had lower beneficiary expenditures and odds of hospitalization. [14] Similarly, O'Malley et al. (2019) found that PCPs with greater comprehensiveness scores (i.e., provided more comprehensive care) had lower beneficiary expenditures and hospitalization rates. [15] Other studies suggested that collaborative care models and home-based visits could reduce hospitalizations. Matzke et al. (2018) found that physician-pharmacist collaborative care-based patient-centered medical home models reduced hospital admissions [16]. Ruiz et al. (2017) showed that some home visit program models significantly reduced hospitalization rates, while Edwards et al. (2017) found that home-based primary care versus usual care was associated with lower odds of admissions for ambulatory care sensitive conditions [17-18].

The MIPS MCC admission measure is consistent with CMS's goal of providing eligible clinicians with actionable data, while at the same time providing patients with a meaningful outcome. CORE expects that sharing measure scores with eligible clinicians, in addition to tying reimbursement payment adjustments to these scores, will strongly encourage eligible clinicians to improve care quality and patient outcomes. Overall, this measure will provide CMS with valuable tools for assessing the performance of outpatient clinicians and groups of clinicians in the MIPS program.

Citations:

Van Cleave JH, Egleston BL, Abbott KM, Hirschman KB, Rao A, Naylor MD. Multiple Chronic Conditions and Hospitalizations Among Recipients of Long-Term Services and Supports. Nurs Res. 2016;65(6):425-434.

Jackson GL, Powers BJ, Chatterjee R, et al. The patient centered medical home. A Systematic Review. Ann Intern Med. 2013;158(3):169-178.

Kern LM, Edwards A, Kaushal R. The Patient-Centered Medical Home and Associations With Health Care Quality and Utilization: A 5-Year Cohort Study. Ann Intern Med. 2016;164(6):395-405.

Maeng DD, Khan N, Tomcavage J, Graf TR, Davis DE, Steele GD. Reduced acute inpatient care was largest savings component of Geisinger Health System's patient-centered medical home. Health Aff (Millwood). 2015;34(4):636-644.

Casalino LP, Pesko MF, Ryan AM, et al. Small primary care physician practices have low rates of preventable hospital admissions. Health Aff (Millwood). 2014;33(9):1680-1688.

Chan DK, Chong R, Basilikas J, Mathie M, Hung WT. Survey of major chronic illnesses and hospital admissions via the emergency department in a randomized older population in Randwick, Australia. Emergency medicine (Fremantle, W.A.). 2002;14(4):387-392.

Glynn LG, Valderas JM, Healy P, et al. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Fam Pract. 2011;28(5):516-523.

Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Aff (Millwood). 2012;31(6):1156-1166.

van Loenen T, van den Berg MJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Fam Pract. 2014;31(5):502-516.

Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. U Am Geriatr Soc. 2008;56(12):2195-2202.

Levine S, Steinman BA, Attaway K, Jung T, Enguidanos S. Home care program for patients at high risk of hospitalization. American Journal of Managed Care. 2012;18(8):e269-276.

Littleford A, Kralik D. Making a difference through integrated community care for older people. Journal of Nursing and Healthcare of Chronic Illness. 2010;2(3):178-186.

Zhang NJ, Wan TT, Rossiter LF, Murawski MM, Patel UB. Evaluation of chronic disease management on outcomes and cost of care for Medicaid beneficiaries. Health policy (Amsterdam, Netherlands). 2008;86(2-3):345-354.

Bazemore, A., et al. (2018). "Higher Primary Care Physician Continuity is Associated With Lower Costs and Hospitalizations." Ann Fam Med. 16(6): 492-497.

O'Malley, A. S., et al. (2019). "New approaches to measuring the comprehensiveness of primary care physicians." Health Serv Res. 54(2): 356-366.

Matzke GR, Moczygemba LR, Williams KJ, Czar MJ, Lee WT. Impact of a pharmacist–physician collaborative care model on patient outcomes and health services utilization. American Journal of Health-System Pharmacy. 2018;75(14):1039-1047

Ruiz S, Snyder LP, Rotondo C, Cross-Barnet C, Colligan EM, Giuriceo K. Innovative Home Visit Models Associated With Reductions In Costs, Hospitalizations, And Emergency Department Use. Health Affairs. 2017;36(3):425-432

Edwards ST, Saha S, Prentice JC, Pizer SD. Preventing Hospitalization with Veterans Affairs Home-Based Primary Care: Which Individuals Benefit Most? Journal of the American Geriatrics Society. 2017;65(8):1676-1683

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.*

Hospital admission rates are an effective marker of ambulatory care quality. Hospital admissions from the outpatient setting reflect a deterioration in patients' clinical status and as such reflect an outcome that is meaningful to both patients and providers. Patients receiving optimal, coordinated high-quality care should use fewer inpatient services than patients receiving fragmented, low-quality care. Thus, high population rates of hospitalization may, at least to some extent, signal poor quality of care or inefficiency in health system performance.

Patients with MCCs are at high risk for hospital admission, often for potentially preventable causes, such as exacerbation of pulmonary disease. [1] Evidence from several Medicare demonstration projects suggests that care coordination results in decreased hospital admission rates among high-risk patients. [2] In addition, studies have shown that the types of ambulatory care clinicians this measure targets (for example, primary care providers and specialists caring for patients with MCCs) can influence admission rates through primary care clinician supply, continuity of care, and patient-centered medical home interventions such as team-based and patient-oriented care. [3-5] More recent studies speak directly to the positive effect that individual providers and group practices can have on lowering patients' hospital visit rates. In particular, they support that comprehensive and continuous care by individual providers can decrease care utilization. [6-7]

Thus, the anticipated net benefits of this unplanned hospital admission measure include, but are not limited to:

- Reduced numbers of hospitalizations and days hospitalized;
- Improved outpatient disease management;
- Reduced rates of complications, including mortality; and
- Cost savings resulting from fewer hospitalizations.

Overall, this measure will provide CMS with a valuable tool for assessing the performance of TINs (individual clinicians and groups of clinicians) in the MIPS program.

Citations

1. Abernathy K, Zhang J, Mauldin P, et al. Acute Care Utilization in Patients With Concurrent Mental Health and Complex Chronic Medical Conditions. J Prim Care Community Health. 2016;7(4):226-233.
2. Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. Health Aff (Millwood). 2012;31(6):1156-1166.
3. van Loenen T, van den Berg MJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Fam Pract. 2014;31(5):502-516.
4. Dale SB, Ghosh A, Peikes DN, et al. Two-Year Costs and Quality in the Comprehensive Primary Care Initiative. N Engl J Med. 2016;374(24):2345-2356.
5. Casalino LP, Pesko MF, Ryan AM, et al. Small primary care physician practices have low rates of preventable hospital admissions. Health Aff (Millwood). 2014;33(9):1680-1688
6. Bazemore, A., et al. (2018). "Higher Primary Care Physician Continuity is Associated With Lower Costs and Hospitalizations." Ann Fam Med. 16(6): 492-497.
7. O´Malley, A. S., et al. (2019). "New approaches to measuring the comprehensiveness of primary care physicians." Health Serv Res. 54(2): 356-366.

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.*

In the calendar year 2018 performance period, a total of 4,659,922 Medicare FFS MCC patients were attributed to 58,435 MIPS-eligible TINs. Acute, unplanned hospital admissions were identified using 2018 Medicare FFS institutional inpatient claims. Overall, across all TINs with at least one attributed MCC patient, RSAAR measure scores ranged from 17.5 to 131.5 per 100 person-years, with a median of 38.7 and an interquartile range of 36.5 to 41.8. The mean RSAAR and standard deviation were 39.5 ± 5.8 admissions per 100 person-years. Generally, similar distributions in measure scores were found across TINs with different provider composition. Below we show the range of RSAAR within each decile for all TINs with at least one attributed MCC patient:

Decile_1// 17.5 - 33.5

Decile_2// 33.5 - 35.7

Decile_3// 35.7 - 37.1

Decile_4// 37.1 - 38.1

Decile_5// 38.1 - 38.7

Decile_6// 38.7 - 39.7

Decile_7// 39.7 - 41.0

Decile_8// 41.0 - 42.8

Decile_9// 42.8 - 46.3

Decile_10// 46.3 - 131.5

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; we provide performance data in 1b.2.

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

Please note that this is a new measure, not a maintenance measure.

Using 2018 Medicare claims data for 4,659,922 Medicare FFS beneficiaries with multiple chronic conditions (MCCs), the final patient-level model includes two social risk factors: the Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index (lowest quartile vs. upper three quartiles) and an area-level measure of specialist physician density (lowest quartile vs. upper three quartiles). In the multivariable model that included both of these social risk factors along with the demographic and clinical risk adjusters, we found relatively modest effects for the social risk factor variables. Their rate ratios and 95% confidence intervals were 1.08 (1.07, 1.08) for the AHRQ SES variable and 1.04 (1.04, 1.05) for the specialist density variable.

The measure does not adjust for dual eligibility (DE). Below we show the distribution of measure scores by TINs stratified by the proportion of patients with the DE social risk factor for all TINs with at least one attributed patient.

Quartile for proportion of DE patient (range of proportion)//Q1 (0.0%- 0.0%)//Q2 (0.2%-8.3%)//Q3 (8.3%-28.5%//Q4 (28.6% - 100.0%)

Number of TINs//17,773//11,578//14,393//14,691

Mean//38.7//38.8//40.5//40.2

Standard Deviation (SD)//3.5//6.5//6.5//6.4

Maximum// 68.5//91.3//104.4//131.5

99th Percentile//50.0//57.7//61.9//62.4

95th Percentile//44.6//50.3//52.2//51.7

90th Percentile//42.5//47.0/48.5//47.4

Upper Quartile//40.3//42.4//43.4//42.4

Median//38.6//38.1//39.4//39.1

Lower Quartile//37.2//34.4//36.4//37.0

10th Percentile//34.9//31.4//33.7//33.9

5th Percentile// 33.3//29.6//31.9//31.6

1st Percentile//29.5//26.1//28.4//26.8

Minimum//20.1//17.5//22.0//18.0

For the 4,044 TINs with at least 15 providers and at least 18 patients, the measure scores by quartile of patients with social risk factors is as follows:

Number of TINs: 118//1,335//1,814//777

Mean: 38.9//39.8//41.9//42.3

Std Dev: 6.1//7.0//7.4//8.9

Maximum: 63.1//91.3//98.7//77.9

99th Percentile: 57.7//59.3//64.6//71.3

95th Percentile: 49.7//51.3//54.4//58.9

90th Percentile: 47.2//48.0//50.6//54.0

Upper Quartile: 41.3//43.6//46.0//46.9

Median: 38.1//39.3//41.1//40.9

Lower Quartile: 34.5//35.0//36.9//36.2

10th Percentile: 32.5//31.7//33.7//32.4

5th Percentile: 30.8//29.6//31.3//30.3

1st Percentile: 28.8//25.9//28.7//25.8

Minimum: 28.5//20.4//23.7//22.2

For more information about the testing of disparities data, please review section 1.8 and section 2b3.3 in the testing attachment.

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

Please see testing described in section 1b.4 above and in sections 1.8 and 2b3.3 of the testing attachment.

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):*

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

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

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.)*

Not applicable.

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_MIPS_MCC_DataDictionary_07302020-637402642885077993.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.

Not applicable. This is a new measure.

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 the number of acute admissions per 100 person-years at risk for admission during the measurement period.

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).*

Outcome Definition

The outcome for this measure is the number of acute, unplanned hospital admissions per 100 person-years at risk for admission during the measurement period.

Time Period

Number of admissions are counted while the patient is considered at risk for an admission during the measurement year.

Excluded Admissions

The numerator (outcome) does not include the following admissions because they do not reflect the quality of care provided by ambulatory care clinicians who are managing the care of MCC patients:

1. Planned hospital admissions;
2. Admissions that occur directly from a skilled nursing facility (SNF) or acute rehabilitation facility;
3. Admissions that occur within a 10-day "buffer period" of time after discharge from a hospital, SNF, or acute rehabilitation facility;
4. Admissions that occur after the patient has entered hospice;
5. Admissions related to complications of procedures or surgeries;
6. Admissions related to accidents or injuries; or
7. Admissions that occur prior to the first visit with the assigned clinician or clinician group.

Clarification regarding the 10-day "buffer period"

The 10-day "buffer period" is a numerator (or outcome) exclusion but it also affects the denominator (person-time at risk); see below in Section S.6 and S.7. The 10-day buffer period (10 days following discharge from a hospital) is a period of transition back to community-based care, and other factors in addition to ambulatory care, including care received in the hospital and post-discharge planning, contribute to the risk of admission; therefore, the measure does not hold clinicians accountable for admissions in this timeframe. This buffer period allows time for patients to be seen within 7 days of discharge as recommended in CMS's Transitional Care Management (TCM) service guidelines and for the ambulatory care provider's care plan to take effect. CMS's TCM service guidelines encourage providers to have a face-to-face visit within 7 days of discharge for Medicare patients with high medical decision complexity.

Identification of Planned Admissions

To identify planned admissions, the measure adopted an algorithm previously developed for CMS's hospital readmission measures, CMS's Planned Readmission Algorithm Version 4.0. [1,2] In brief, the algorithm uses the procedure codes and principal discharge diagnosis code on each hospital claim to identify admissions that are typically planned. A few specific, limited types of care are always considered planned (for example, major organ transplant, rehabilitation, and maintenance chemotherapy). Otherwise, a planned admission is defined as a non-acute admission for a scheduled procedure (for example, total hip replacement or cholecystectomy). Admissions for an acute illness are never considered planned. For specific codes included in the planned admission algorithm please see Tables PAA1-PAA4 with the codes for the CMS Planned Admission Algorithm in the accompanying data dictionary.

Identification of admissions that occur directly from an SNF or acute rehabilitation facility

Claims for SNF and acute rehabilitation facility stays, which help determine the outcome definition, were obtained using CMS's Integrated Data Repository (IDR) and Medicare Provider Analysis and Review (MedPAR) files, respectively.

Identification of admissions that occur after the patient has entered hospice

The status of enrollment in Medicare Parts A and B and Medicare's hospice benefit for the measurement year and the year prior were obtained from the CMS Medicare Enrollment Database.

Identification of admissions related to complications of procedures or surgeries (including small bowel obstruction), and accidents or injuries

Using the Agency for Healthcare Research and Quality's (AHRQ's) Clinical Classifications Software (CCS), which clusters diagnoses into clinically meaningful categories using International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes, we exclude from the outcome admissions related to the following 23 CCS categories. For specific ICD codes included, please refer to AHRQ's CCS Version 2019.1, Fiscal Year 2020.

a) Complications of procedures or surgeries
    1. 145: Intestinal obstruction without hernia
    2. 237: Complication of device; implant or graft
    3. 238: Complications of surgical procedures or medical care
    4. 257: Other aftercare

b) Accidents or injuries
    5. 2601 E Codes: Cut/pierce
    6. 2602 E Codes: Drowning/submersion
    7. 2604 E Codes: Fire/burn
    8. 2605 E Codes: Firearm
    9. 2606 E Codes: Machinery
    10. 2607 E Codes: Motor vehicle traffic (MVT)
    11. 2608 E Codes: Pedal cyclist; not MVT
    12. 2609 E Codes: Pedestrian; not MVT
    13. 2610 E Codes: Transport; not MVT
    14. 2611 E Codes: Natural/environment
    15. 2612 E Codes: Overexertion
    16. 2613 E Codes: Poisoning
    17. 2614 E Codes: Struck by; against
    18. 2615 E Codes: Suffocation
    19. 2616 E Codes: Adverse effects of medical care
    20. 2618 E Codes: Other specified and classifiable
    21. 2619 E Codes: Other specified; NEC
    22. 2620 E Codes: Unspecified
    23. 2621 E Codes: Place of occurrence

Citations

1. Yale New Haven Health Services Corporation – Center for Outcomes Research & Evaluation (YNHHSC/CORE). 2018 All-Cause Hospital Wide Measure Updates and Specifications Report - Hospital-Level 30-Day Risk-Standardized Readmission Measure – Version 7.0. Centers for Medicare & Medicaid Services; March 2018.

2. Horwitz L, Grady J, Cohen D, et al. Development and validation of an algorithm to identify planned readmissions from claims data. Journal of Hospital Medicine. Oct 2015;10(10):670-677.

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

Patients included in the measure (target patient population)

The target patient population for the outcome includes Medicare FFS patients aged 65 years and older with multiple chronic conditions (MCCs).

Provider types included for measurement

- Primary care providers (PCPs): CMS designates PCPs as physicians who practice internal medicine, family medicine, general medicine, or geriatric medicine, and non-physician providers, including nurse practitioners, certified clinical nurse specialists, and physician assistants.
- Relevant specialists: Specialists covered by the measure are limited to those who provide overall coordination of care for patients with MCCs and who manage the chronic diseases that put the MCCs patients in the measure at risk of admission. These specialists were chosen with input from our Technical Expert Panel (TEP) and include cardiologists, pulmonologists, nephrologists, neurologists, endocrinologists, and hematologists/oncologists. However, as indicated below and in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.

Patient attribution

We begin by assigning each patient to the clinician most responsible for the patient's care. The patient can be assigned to a PCP, a relevant specialist, or can be left unassigned.

- A patient who is eligible for attribution can be assigned to a relevant specialist only if the specialist has been identified as "dominant". A specialist is considered "dominant" if they have two or more visits with the patient, as well as at least two more visits than any PCP or other relevant specialist. For example, if a patient saw a cardiologist four times in the measurement year, a PCP twice, and a nephrologist twice, the patient would be assigned to the cardiologist, having met the definition of "dominant" specialist. Note: Hematologists and oncologists are considered relevant specialists as they could be expected to manage MCCs patients' care, especially during periods of acute cancer treatment. However, as indicated below in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.
- There are two scenarios where a patient can be assigned to a PCP. First, the patient must have seen at least one PCP. The patient will then be assigned to the PCP with the highest number of visits as long as there is no relevant specialist who is considered "dominant." Second, if the patient has had more than one visit with a relevant specialist but no "dominant" specialist has been identified, and has two or more visits with a PCP, they will be assigned to that PCP.
- Finally, the patient will be unassigned if they only saw non-relevant specialists, if the patient has not seen a PCP and no "dominant" specialist can be identified, or if the patient has not had more than one visit with any individual PCP.

Patients are then assigned at the Taxpayer Identification Number (TIN) level, which includes solo clinicians and groups of clinicians who have chosen to report their quality under a common TIN.

- At the TIN level, patients are first assigned to the clinician (unique National Provider Identifier (NPI)/TIN combination since a given provider can be affiliated with more than one TIN) most responsible for their care (using the algorithm for individual clinician-level attribution above) and then patients "follow" their clinician to the TIN designated by the clinician. Patients unassigned at the individual clinician level continue to be unassigned at the TIN level.

(Note that an alternative attribution approach was considered and assessed as described in section 2b.3.11 of the testing attachment and in Appendix C of the attached methodology report.)

Person-time at risk

Persons are considered at risk for hospital admission if they are alive, enrolled in FFS Medicare, and not in the hospital during the measurement period. In addition to time spent in the hospital, we also exclude from at-risk time: 1) time spent in a SNF or acute rehabilitation facility; 2) the time within 10 days following discharge from a hospital, SNF, or acute rehabilitation facility; and 3) time after entering hospice care.

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).*

Patients included in the measure (target patient population)

The cohort, or group of patients included in the measure, is comprised of patients whose combinations of chronic conditions put them at high risk of admission and whose admission rates could be lowered through better care. This definition reflects NQF's "Multiple Chronic Conditions Measurement Framework," which defines patients with MCCs as people "having two or more concurrent chronic conditions that … act together to significantly increase the complexity of management, and affect functional roles and health outcomes, compromise life expectancy, or hinder self-management." [1]

The specific inclusion criteria are as follows.

1. Patient is alive at the start of the measurement period and has two or more of nine chronic condition disease groups in the year prior to the measurement period.

Chronic conditions, except for diabetes, are defined using CMS's Chronic Conditions Data Warehouse (CCW). For diabetes, we used the diabetes cohort definition from the Accountable Care Organization (ACO) diabetes admission measure developed by CORE (v2018a ACO-36) as opposed to the definition used in CCW, which includes diagnoses for secondary and drug-induced diabetic conditions that are not the focus of the MIPS MCCs admission measure. See Table 1 in the accompanying data dictionary for the specific codes used to define the nine cohort-qualifying conditions.

1. Acute myocardial infarction (AMI),
2. Alzheimer's disease and related disorders or senile dementia,
3. Atrial fibrillation,
4. Chronic kidney disease (CKD),
5. Chronic obstructive pulmonary disease (COPD) or asthma,
6. Depression,

7. Diabetes,
8. Heart failure, and
9. Stroke or transient ischemic attack (TIA).

Rationale: As noted above, this definition of MCCs is consistent with NQF's "Multiple Chronic Conditions Measurement Framework." The specific list of chronic conditions, except for diabetes, is the same as is used in the MCCs admission measure for ACOs (ACO-38) currently implemented the Medicare Shared Savings Program. This measure has been vetted nationally and published in the literature. [2] In brief, it reflects the chronic conditions that most increased risk of admission. In adapting the ACO measure for the MIPS setting, we added diabetes as a cohort-qualifying condition based on input from our TEP and further guidance from CMS. In addition, the inclusion of diabetes acknowledges the complexity that diabetes introduces to caring for patients with MCCs.

2. Patient is aged =65 years at the start of the year prior to the measurement period.

Rationale: Younger Medicare patients represent a distinct population with dissimilar characteristics and outcomes. Additionally, these patients tend to cluster among certain providers. These factors make risk adjustment difficult.

3. Patient is a Medicare FFS beneficiary with continuous enrollment in Medicare Parts A and B during the year prior to the measurement period.

Rationale: Enrollment is necessary to provide clinical information for cohort identification

and risk adjustment.

Provider types included for measurement

Because we use the outcome of acute, unplanned admissions to assess quality, we limit the clinicians covered by the measure to two categories of providers for whom this outcome reflects care quality. This includes primary care providers (PCPs) and a subset of specialists who manage the care of MCCs patients.

Primary Care Providers - CMS designates PCPs as physicians who practice:

1. Internal medicine,
2. Family medicine,
3. General medicine, or
4. Geriatric medicine; and

The following non-physician clinicians:

1. Nurse practitioners,
2. Certified clinical nurse specialists, and
3. Physician assistants. [3]

Relevant specialists - Based on input from the TEP, specialists covered by the measure are limited to those who plausibly provide overall coordination of care for patients with MCCs and who manage the chronic diseases that put the MCCs patients in the measure at risk of admission. These "relevant" specialists, defined using the Medicare Provider Specialty Codes (see Table 4 in the accompanying data dictionary), are:

1. Cardiologists,
2. Pulmonologists,

3. Nephrologists,
4. Neurologists,
5. Endocrinologists, and
6. Hematologists/oncologists.

Note: Hematologists and oncologists are considered relevant specialists as they could be expected to manage MCCs patients' care, especially during periods of acute cancer treatment. However, as indicated below in Section S.9, the measure is not designed to assess the quality of care of cancer specialists who are actively managing cancer patients, and thus patients attributed to hematologists and oncologists are excluded from the measure.

Patient attribution

As noted in field Section S.6., we use a visit-based algorithm to assign MCCs patients to the individual clinician most responsible for their care. The attribution approach uses the plurality of Evaluation and Management (E&M) visits. (Please see Table 3 in the accompanying data dictionary for specific codes.) Focusing on visits over charges when assigning responsibility acknowledges the importance of provider interaction with the patient in establishing accountability for outcomes. In most instances, the provider with the most visits is a PCP.

Citations

1. National Quality Forum. Multiple Chronic Conditions Measurement Framework. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=71227. Accessed February 20, 2019.
2. Drye EE, Altaf FK, Lipska KJ, et al. Defining Multiple Chronic Conditions for Quality Measurement. Med Care. 2018;56(2):193-201.
3. Centers for Medicare & Medicaid Services. Medicare Claims Processing Manual Chapter 4 - Part B Hospital (Including Inpatient Hospital Part B and OPPS) (section 250.12.1). https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c04.pdf. Accessed February 20, 2019.

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

We exclude patients from the cohort for these reasons:

1. Patients without continuous enrollment in Medicare Part A or B during the measurement period.
2. Patients enrolled in hospice at any time during the year prior to the measurement year or at start of the measurement year.
3. Patients with no E&M visit to a MIPS eligible clinician.
4. Patients assigned to clinicians who do not participate in the QPP on the MIPS track.
5. Patients attributed to hematologists and oncologists.
6. Patients not at risk for hospitalization during the measurement 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.)*

The rationale for each exclusion is provided below:

1. Patients without continuous enrollment in Medicare Part A or B during the measurement period.

Rationale: The measure excludes these patients to ensure full data availability for outcome assessment and attribution.

2. Patients enrolled in hospice at any time during the year prior to the measurement year or at start of the measurement year.

Rationale: The measure excludes these patients even though once a patient enters hospice care, a goal of care is to prevent the need for hospital care. However, ambulatory care providers may have relatively little influence on end-of-life care once a patient is enrolled in hospice and served by a hospice team.

3. Patients with no E&M visit to a MIPS eligible clinician.

Rationale: The measure excludes these patients because they could not be attributed to a provider using the visit-based attribution algorithm (see Section S.6 for details).

4. Patients assigned to clinicians who do not participate in the QPP on the MIPS track.

Rationale: These patients are excluded because the clinicians to whom they are assigned do not participate in MIPS.

5. Patients attributed to hematologists and oncologists.

Rationale: The outcomes for patients who are predominantly cared for by hematologists and oncologists, including patients actively being managed for cancer, are unlikely to reflect the quality of care provided by primary care provider (PCP) or other relevant specialists. The aim of this measure is not to assess the quality of care during such instances of active cancer treatment. Excluding patients assigned to hematologists and oncologists takes out of the measure patients who are being actively treated for cancer during the measurement period but retains in the measure patients with MCCs who have a history of cancer or are occasionally being seen by a cancer specialist for follow-up.

6. Patients not at risk for hospitalization during the measurement year.

Rationale: The outcomes for these patients cannot be assessed as they are not at risk. For example, if the first attributed visit occurred after the patient has entered hospice, the patient would not have any time at risk and would thus be excluded. See section 2.4.3 of the attached methodology report for methods used to calculate person-time at risk.

Clarification of 10-day buffer period:

The 10-day "buffer period" is a numerator (or outcome) exclusion (see section S.5) but it also affects the denominator (person-time at risk). Persons are considered at risk for hospital admission if they are alive, enrolled in FFS Medicare, and not in the hospital during the measurement period. In addition to time spent in the hospital, we also exclude from at-risk time:

1) time spent in a SNF or acute rehabilitation facility;
2) the time within 10 days following discharge from a hospital, SNF, or acute rehabilitation facility; and
3) time after entering hospice care. Note that the patient is not removed from the denominator, we are just subtracting the 10-days of person-time.

The 10-day buffer period (10 days following discharge from a hospital) is a period of transition back to community-based care, and other factors in addition to ambulatory care, including care received in the hospital and post-discharge planning, contribute to the risk of admission; therefore, the measure does not hold clinicians accountable for admissions in this timeframe. This buffer period allows time for

patients to be seen within 7 days of discharge as recommended in CMS's Transitional Care Management (TCM) service guidelines and for the ambulatory care provider's care plan to take effect. CMS's TCM service guidelines encourage providers to have a face-to-face visit within 7 days of discharge for Medicare patients with high medical decision complexity.

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; this measure is not stratified.

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.)*

We begin by identifying the cohort of patients with MCCs by applying the inclusion/exclusion criteria. We then use the attribution algorithm to assign patients to TINs. Patients are assigned to the individual clinician most responsible for their care, and then subsequently to the TIN designated by the clinician, using our visit-based attribution algorithm. Attribution is done in the measurement period and only patients assigned to a MIPS-eligible clinician will be included in the measure score calculation. The number of admissions and time at risk in the measurement period are then calculated for each patient based on our measure specifications. The measure is risk-adjusted for demographic, clinical, and social risk factors. For the score calculation, the measure uses a hierarchical (two-level) statistical model that accounts for the clustering of patients within MIPS providers and accommodates the varying patient sample sizes of different providers. The measure uses a negative binomial with linear variance (NB-1) model since the measure's outcome is a count of the number of admissions for MCCs patients during the measurement period. The first level of the model adjusts for patient factors. The relationship between patient risk factors and the outcome of admissions is determined based on all patients attributed to MIPS-eligible clinicians. Therefore, the "expected" number of admissions (described below) for each provider is based on the performance of all eligible MIPS providers nationwide.

The second level of the model estimates a random-intercept term that reflects the provider's contribution to admission risk, based on their actual admission rate, the performance of other providers, their case mix, and their sample size.

The measure score is a risk-standardized acute admission rate (RSAAR), calculated as the ratio of the number of predicted admissions to the number of expected admissions multiplied by the crude national rate. The predicted to expected ratio of admissions is analogous to an observed over expected ratio, but the numerator accounts for clustering, sample-size variation, and provider-specific performance. The expected number of admissions is calculated based on the provider's case mix and average intercept among all MIPS providers. The predicted number of admissions is calculated based on the provider's case mix and the estimated provider-specific random intercept term. We multiply the predicted to expected ratio for each provider by a constant – the crude rate of acute, unplanned admissions among all MIPS providers – for ease of interpretation.

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.

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 survey or instrument.

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.

Medicare administrative claims and enrollment data, American Community Survey, Area Health Resource Files; dates vary; see Section 1.7 of the testing attachment for details.

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)*

Clinician : Group/Practice

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

Outpatient Services

If other:

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

N/A; this measure is not a composite.

2. Validity – See attached Measure Testing Submission Form

NQF_MIPS_MCC_TestingForm_110520-637418981680116771.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.

No

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.

No

2.3 *For maintenance of endorsement*

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** N/A

Measure Title: Clinician Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System

Date of Submission: **11/6/2020**

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

Instructions

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

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

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

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

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

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: All data sources are described below in section 1.2	☒ other: All data sources are described below in section 1.2

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 datasets, dates, and number of measured entities used in each type of testing were as follows:

For measure development and initial testing, we used Medicare Fee-for-Service (FFS) administrative claims data (Parts A and B) and the Medicare Enrollment Database (EDB) from calendar years 2013 through 2015, as well as the additional datasets specified below.

a. Datasets used to define the cohort:

- The status of enrollment in Medicare Parts A and B and Medicare's hospice benefit for 2014-2015 were obtained from the CMS Medicare EDB and used to define cohort eligibility. Chronic conditions, except for diabetes, were defined using CMS's Chronic Conditions Data Warehouse (CCW). For diabetes, we used the diabetes cohort definition from the ACO diabetes admission measure currently in use by CMS (NQF #2887) as opposed to the definition used in CCW, which includes diagnoses for secondary and drug-induced diabetic conditions that are not the focus of the MIPS MCC admission measure.
- Our data predated the MIPS program; therefore, the cohort of beneficiaries assigned to MIPS-eligible clinicians was approximated by including only those patients attributed to Value Modifier-eligible providers or non-risk-bearing Shared Savings Program ACOs in 2015 – the providers most likely to be eligible to participate in the QPP under MIPS. The patient-provider assignments for 2015 were provided by CMS's Value Modifier program contractor.

b. Datasets used to capture the outcome:

- We assessed provider performance in calendar year 2015 (referred to as the measurement year). The outcome of acute, unplanned hospital admissions was identified using 2015 Medicare FFS institutional inpatient claims. Information on skilled nursing facility (SNF) and acute rehabilitation

facility stays as well as hospice entry, which factor into the outcome definition, was obtained using CMS's Integrated Data Repository (IDR), Medicare Provider Analysis and Review (MedPAR) files, and Medicare Enrollment Database (EDB), respectively. Information on prior year's visit to attributed clinician, used in determining whether to exclude some admissions in the measurement year, was obtained from 2014 Medicare FFS non-institutional carrier claims.

c. Datasets used for attribution:

- Outpatient Evaluation and Management (E&M) visits during the measurement year, identified using the 2015 Medicare FFS non-institutional carrier claims, were used to attribute patients to providers.

d. Datasets used to identify risk adjustment factors:

- Clinical comorbidities: 2014 Medicare FFS institutional inpatient and outpatient claims, and non-institutional carrier claims.
- Frailty indicators: 2014 durable medical equipment (DME) claims.
- Original reason for Medicare entitlement: Medicare EDB.
- Social risk factors: The social risk factors for analysis and the data sources used to define them were as follows:
  1) Medicare/Medicaid dual-eligibility status: 2014-2015 Medicare Master Beneficiary Summary File (MBSF).
  2) Agency for Healthcare Research Quality (AHRQ) Socioeconomic Status (SES) Index: 2009-2013 American Community Survey (ACS).
  3) Rural residence: 2014 United States Department of Agriculture Economic Research Service.
  4) Primary care provider (PCP) and physician-specialist density variables: 2017-2018 Area Health Resources File (AHRF).

For updated testing of the measure, we used updated data as follows:

*Medicare administrative claims and enrollment data from calendar years 2017 and 2018

*2013-2017 American Community Survey, and 2018-2019 Area Health Resource File

*For the cohort, we used claims from the Integrated Data Repository (IDR)

*For MIPS eligibility, we used the MIPS-eligibility file from the QPP CMS contractor

*For Rehabilitation stays, we used data from the IDR

See Section 1.7 for details.

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

The dates of the data vary by testing type. More information is provided in Section 1.7.

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☐ 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*)

Taxpayer Identification Number (TIN)-based groups/practices, which can include individual clinicians, in the Merit-based Incentive Payment System are the measured entities. The number of measured entities 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 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.

Dataset	Description of Dataset	Use and Section in the Testing Attachment
Dataset #1: 2015 Medicare MCC Full Sample Dataset #1a: Development Sample Dataset #1b: Validation Sample	For measure development and testing, we used Medicare Fee-for-Service (FFS) administrative claims data (Parts A and B) and the Medicare Enrollment Database (EDB) from calendar years 2013 through 2015 (see Section 1.2). The primary dataset used for testing is referred to as the "2015 Medicare MCC Full Sample." **2015 Medicare MCC Full Sample** - Dates: January 1, 2013 – December 31, 2015 - Number of clinician groups/practices (TINs): 64,086 - Number of patients: 4,937,865 - Patient demographic and clinical characteristics: The sample was 57.9% female and had an average age of 79.0 ±7.9 years. **Development and Validation Samples** For development and testing of the patient-level model, we randomly split the 2015 Medicare MCC Full Sample into Development and Validation samples. The Development Sample included a random 50% sample, and the Validation Sample included the remaining 50% of MCC patients not selected into the Development Sample. *Development Sample* - Dates: January 1, 2013 – December 31, 2015 - Number of patients: 2,468,933 - Patient demographic and clinical characteristics: The sample was 57.9% female and had an average age of 79.0 ±7.9 years. *Validation Sample* - Dates: January 1, 2013 – December 3, 2015 - Number of patients: 2,468,932 - Patient demographic and clinical characteristics: The sample was 57.9% female and had an average age of 79.0 (±7.9 years).	• Measure development, initial testing, and face validity testing (Section 2b1.2) • Selection of demographic and clinical variables (Section 2b3.1.1) ○ • Model performance testing (Section 2b3.5) ○ ○

Dataset	Description of Dataset	Use and Section in the Testing Attachment
Dataset #2: ICD-10 Testing Dataset	For updated testing we used the **ICD-10 Testing Dataset:** - Dates: January 1, 2018 – December 31, 2018 - Number of patients: 4,659,922 - Number of clinician groups (TINs) with at least one patient with MCCs: 58,435 - Number of clinician groups (TINs) with at least 15 providers and 18 patients with MCCs: 4,044 - Patient demographic and clinical characteristics: The sample was 57.5% female and had an average age of 79.0 ±8.0 years.	• Reliability testing (Section 2a2.2), • Exclusions (Section 2b2.2) • Social risk factor testing (Section 2b3.4b) • Model performance (Section 2b3.5) • Meaningful Differences (Section 2b4) • Missing data (Section 2b6)

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 developed and used the conceptual model described in Section 2b3.3a below to select potential social risk factors. Based on our conceptual model, we analyzed the following 5 social risk factors:

1. Medicare-Medicaid dual-eligibility status

Direct measures of Medicare beneficiaries' socioeconomic status (SES) are not available. We included the readily available and widely used dual-eligibility status variable as it is a marker of low income and assets.

2. Agency for Healthcare Research and Quality (AHRQ) SES Index

The AHRQ SES Index is a widely used variable that summarizes area-level measures of employment, income, education, and housing. Each of the index components is available at the census block level, which we then used to link to patient's residence using 9-digit ZIP code. Census variables were found in the American Community Survey. The AHRQ SES index score summarizes the information from the following 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.

3. Place of residence (rurality)

We categorized beneficiaries' place of residence in terms of rurality, given its implications for timely receipt of care and concerns that individuals in more rural areas may suffer delays due to longer travel distance and time and relative lack of providers.

4-5. Provider density

To more fully and directly characterize access to care, we additionally included two measures of provider density: 1) PCP density and 2) physician-specialist density.

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 considered signal-to-noise analysis as a measure of reliability when evaluating the MIPS MCC admission measure. The variation between entities ('signal') comprises the total variation ('noise' and 'signal') in the outcome. This is because the reliability of any one TIN's measure score will vary depending on the number of patients. Entities with higher volume will tend to have more reliable scores, while those with lower volume will tend to have less reliable scores. We used the formula for signal-to-noise reliability presented by Adams et al. and the formula for intraclass correlation coefficient (ICC) presented by Nakagawa et al. to calculate individual clinician-level and TIN-level reliability scores.[1,2] To estimate the overall signal and noise, we first calculated the ICC for the provider entity (TIN) j using the estimates of between-entity variance 2, dispersion parameter ,, and mean of outcome , from a hierarchical generalized linear model (HGLM). The formula appropriate for the NB-1 model is ICCj = 2/(2 + ln(1+/).

We then used the equation:

$$Rj = njICCj/(1+(nj-1)ICCj)$$

where nj is the number of observations for each entity, to calculate the reliability of each entity measurement. Rj can range from 0 (less than chance agreement) to 1.0 (perfect agreement).

In addition, we determined the minimum number of patients required to achieve reliability values of 0.4, 0.5, 0.6, and 0.7, using the equation:

$$nj = Rj*(1-ICCj)/(ICCj*(1-Rj))$$

where Rj = 0.4, 0.5, 0.6, or 0.7. We reported the distribution of Rj over all entities and for those meeting the different volume requirement nj.

Citations

1. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician Cost Profiling — Reliability and Risk of Misclassification. New England Journal of Medicine. 2010;362(11):1014-1021.
2. Nakagawa S, Johnson PCD, Schielzeth H. The coefficient of determination R(2) and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface. 2017;14(134).

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*)

Among all MIPS TINs with at least one attributed MCC patient (n=58,435 TINs), mean and median signal-to-noise reliability for the MIPS MCC measure was 0.453 and 0.451, respectively (range 0.038-0.999, Interquartile Range (IQR) 0.190-0.694) (Table 1).

Since the measure is intended to be used for clinician groups, we also calculated reliability for the measure when applied to TINs with more than one clinician. To define practice group size, we calculated a count of NPIs associated with each TIN. The NPI count per TIN was based on any NPI associated with that TIN, irrespective of NPI specialty or MIPS eligibility.

Using a group size threshold of >15 clinicians per group, mean and median reliability for 6,326 TINs was 0.580 and 0.648, respectively (range 0.038-0.999, IQR 0.238-0.926) (Table 1).

We then examined the distribution of mean and median reliabilities across patient volume for these clinician groups and identified a case minimum of 18 MCC patients per clinician group as providing adequate reliability. Using this case minimum and the group size threshold of >15 clinicians per group, mean and median reliability for 4,044 TINs was 0.809 and 0.873, respectively (range 0.413-0.999, IQR 0.683-0.961) (Table 1).

Table 1: Signal-to-noise reliability

TINs	Sample Size Threshold	Number of TINs	Number of MCC Patients	Median Reliability	Mean Reliability	Range	IQR (25^{th}-75^{th})
All TINs	None	58,435	4,659,922	0.451	0.453	0.038-0.999	0.190-0.694
TINs with >15 clinicians	None	6,326	2,408,621	0.648	0.580	0.038-0.999	0.238-0.926
TINs with >15 clinicians	At least 18 MCC patients	4,044	2,394,787	0.873	0.809	0.413-0.999	0.683-0.961

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?*)

The MIPS MCC measure was originally developed for MIPS clinicians and clinician groups. However, during the review of the MIPS MCC measure at the NQF's Measure Applications Partnership (MAP), the committee expressed support for the measure's concept but recommended that the measure be defined for clinician groups only. This recommendation was partly driven by reliability results and partly by concerns that individual clinicians may lack the necessary resources and structural supports to

effectively reduce the risk of admissions among their MCC patients compared with larger groups of clinicians.

Consistent with the recommendation made by the MAP, we provide testing results for the measure using a practice size of greater than 15 providers per group and with a case minimum of at least 18 patients with MCCs. For this subset of TINs, the median reliability is 0.873 which is considered high [1, 2]. A quarter of clinician groups have scores with reliability between 0.4 and 0.7, one-half have scores between 0.7 and about 0.9, and a quarter have scores that well exceed reliability of 0.9.

Citations

1. Yu H, Mehrotra A, Adams J. Reliability of utilization measures for primary care physician profiling. Healthc (Amst). 2013;1(1-2):22-29. doi:10.1016/j.hjdsi.2013.04.002.
2. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician Cost Profiling — Reliability and Risk of Misclassification. New England Journal of Medicine. 2010;362(11):1014-1021.

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*)

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)*

Measure validity is demonstrated through assessment of measure face validity by a technical expert panel (TEP) and use of established measure development guidelines. [1-6]

Validity as Assessed by External Groups and TEP

Throughout the measure development process, we obtained expert and stakeholder input through consulting our national Technical Expert Panel (TEP) and holding a 4-week public comment period.

In addition to the clinical consultations and in alignment with CMS Measures Management System (MMS) guidance, 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 clinicians, patients, and individuals with expertise in quality improvement and performance measurement. We held five structured TEP conference calls consisting of a presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We made modifications to the measure features based on TEP feedback following each meeting.

TEP members were asked to assess the face validity of the final measure specification by confidentially reporting their agreement with following two statements.

"The risk-standardized acute admission rates obtained from the MIPS MCC admission measure as specified:

1) Can be used to distinguish good from poor quality of care provided to MCC patients by TINs reporting under MIPS."
2) Will provide TINs reporting under MIPS with information that can be used to improve their quality of care for MCC patients."

TEP members were asked to report their level of agreement with each statement on a 6-point Likert scale, ranging from "strongly agree" to "strongly disagree;" and to provide rationale for their rating. We also solicited input from the CMS Quality Payment Program (QPP) Clinician Champions specifically around the attribution approach. A list of TEP members is available in the attached methodology report.

Finally, following completion of the preliminary measure, we solicited public comment on the measure through the CMS site: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html.

Validity Indicated by Established Measure Development Guidelines

We developed this measure in consultation with national guidelines for publicly reported outcome measures, with input from outside experts and the public. The measure is consistent with the technical approach to outcome measurement set forth in NQF guidance for outcome measures [7], CMS MMS guidance, and guidance articulated in the American Heart Association scientific statement entitled, "Standards for Statistical Models Used for Public Reporting of Health Outcomes." [8]

Citations

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 Apr 4; 113(13):1683-92.
2. Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. Circulation: Cardiovascular Quality and Outcomes. 2011 Mar 1; 4(2):243-52.
3. 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 Apr 4; 113(13):1693-701.
4. 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: Cardiovascular Quality and Outcomes. 2008 Sep; 1(1):29-37.
5. Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. Public Library of Science (PLoS) ONE. 2011 Apr 12; 6(4):e17401.
6. Lindenauer PK, Normand SL, Drye EE, et al. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. Journal of Hospital Medicine. 2011 Mar; 6(3):142-50.
7. National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report. Available at: http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed June 7, 2017.

8. 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. 2006; 113(3):456-462.

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

Of 17 TEP members who were active through the end of the project, 12 responded. Their responses are reported in the table below.

Frequency of Ratings of Agreement

Question 1: The risk-standardized acute admission rates obtained from the MCC measure as specified can be used to distinguish good from poor quality of care provided to MCC patients by TINs reporting under MIPS.

Rating (# of Responses)

Strongly disagree (1)

Moderately disagree (1)

Somewhat disagree (0)

Somewhat agree (5)

Moderately agree (5)

Strongly agree (0)

Question 2: The risk-standardized acute admission rates obtained from the MCC measure as specified will provide TINs reporting under MIPS with information that can be used to improve their quality of care for MCC patients.

Rating (# of Responses)

Strongly disagree (0)

Moderately disagree (1)

Somewhat disagree (2)

Somewhat agree (3)

Moderately agree (5)

Strongly agree (1)

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?)

The majority of the respondents, 10/12 or 83%, agreed (somewhat or moderately) that the MIPS MCC admission measure can be used to distinguish good from poor quality of care. The majority of the respondents, 9/12 or 75%, agreed (somewhat, moderately, or strongly) that the MIPS MCC admission measure scores (risk-standardized acute admission rates or RSAARs) will provide MIPS TINs with information that could be used to improve the quality of care for MCC patients.

Three of the 12 respondents somewhat, moderately, or strongly disagreed with one or both statements. Of these:

One TEP member noted the measure would not be actionable unless CMS provided patient-level data alongside the overall measure score (RSAAR) to TINs.

The second TEP member was concerned that the measure does not include any element of patient responsibility given challenges patients may encounter in managing their heart failure.

The third TEP member questioned whether lower admission rates are a signal of higher-quality care. Further, the member noted the impacts of risk adjustment and of potential omission of risk variables are unclear. Regarding risk adjustment, the TEP member posited the reliance on diagnosis coding for clinical risk adjustment may favor better-resourced practices that could be more sophisticated in their approaches to coding. The TEP member concluded the risks or harms of deploying the measure may outweigh the benefits.

Overall, the survey indicates support of the validity and usability of the measure.

2b2. EXCLUSIONS ANALYSIS

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

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

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain the impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion. Rationales for the exclusions are detailed in data field S.7 of the Submission Form.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded)*

Following the application of the measure's inclusion criteria, 7,975,506 patients were in the cohort. After applying the exclusions listed below (patients may have more than one exclusion, so the categories are not mutually exclusive), there were 4,659,922 patients remaining in the cohort, attributed to 58,435 TINs with at least one patient with MCCs.

Exclusion	Number of Beneficiaries Excluded	Percent of Beneficiaries Excluded
Patients without continuous enrollment in Medicare Part A or B during the measurement period.	616,406	7.73%
Patient enrolled in hospice at any time during the year prior to the measurement year or at start of the measurement year.	159,857	2.00%
Patients with no E&M visit to a MIPS eligible clinician	226,752	3.00%
Patients assigned to clinicians who do not participate in the QPP on the MIPS track.	818,113	15.7%

Exclusion	Number of Beneficiaries Excluded	Percent of Beneficiaries Excluded
Patients attributed to hematologists and oncologists.	905,452	11.4%
Patients not at risk for hospitalization during the measurement year.	50,166	0.7%

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*)

The exclusions applied to the MIPS MCC measure cohort are required to produce valid, reliable, and fair measure scores.

The rationale for each exclusion is outlined below.

About 8% of the initial index cohort, were excluded from the measure because they were not fully covered under Medicare Parts A and B during the measurement period (2018). These patients were excluded so that full data availability for outcome assessment and attribution was ensured.

About 2.0% of the initial index cohort were excluded from the measure because they were in hospice during the year prior to the measurement year or at the start of the measurement year. The measure excludes these patients even though once a patient enters hospice care, a goal of care is to prevent the need for hospital care. However, ambulatory care providers may have relatively little influence on end-of-life care once a patient is enrolled in hospice and served by a hospice team.

3.0% of the initial index cohort, were excluded from the measure because they did not have any outpatient Evaluation & Management (E&M) visits to a MIPS Medicare provider in 2018 and therefore could not be attributed to a provider. The attribution approach uses the plurality of E&M visits. Focusing on visits over charges when assigning responsibility acknowledges the importance of provider interaction with the patient in establishing accountability for outcomes.

15.7% of patients were excluded after attribution because they were assigned to risk-bearing ACOs likely participating in the Quality Payment Program as advanced APMs. These patients are excluded because the clinicians to whom they are assigned do not participate in MIPS.

11.4% of patients were excluded after attribution because they were assigned to hematologists or oncologists. The outcomes for patients who are predominantly cared for by hematologists and oncologists, including patients actively managed for cancer, are not likely to reflect the quality of care provided by primary care providers or other relevant specialists.

0.7% of patients were excluded after attribution because they were not at risk for hospitalization at any time in 2015. The outcomes for these patients cannot be assessed because they are not at risk.

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

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

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 49 risk adjustment variables.**

☐ **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.

The final patient-level risk-adjustment model included 49 risk-adjustment variables, including 47 demographic and clinical variables and 2 social risk factors. We used a negative binomial regression model with linear variance (NB-1) to risk adjust the measure. See data dictionary Tables 5-8 in the accompanying Excel workbook for specific codes and definitions for each risk variable; the prevalence of each risk variable and the associated rate ratios are shown in the table at the end of this section.

Details of the risk model, including equations, are shown below in this section, as well as in the technical report, in Appendix E1, which is attached to this application.

Final multivariable risk-adjustment model: demographic, clinical, and social risk factors (dataset: 2015 Medicare MCC Full Sample)

Demographic

- Age

Nine chronic disease groups defined using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes

- Acute myocardial infarction
- Alzheimer's disease and related disorders or senile dementia
- Atrial fibrillation
- Chronic kidney disease
- Chronic obstructive pulmonary disease or asthma
- Depression
- Diabetes
- Heart failure
- Stroke or transient ischemic attack

Clinical comorbidities defined using Condition Categories (CCs) or ICD-9-CM codes

- Dialysis status
- Respiratory failure
- Liver disease
- Pneumonia
- Septicemia/shock
- Marked disability/frailty
- Hematologic/al diseases
- Advanced cancer
- Infectious and immune disorders
- Severe cognitive impairment

- Major organ transplant status
- Pulmonary heart disease
- Cardiomyopathy
- Gastrointestinal disease
- Iron deficiency anemia
- Ischemic heart disease except AMI
- Other lung disorders
- Vascular or circulatory disease
- Other significant endocrine disorders
- Other disabilities and paralysis
- Substance abuse
- Other neurologic disorders
- Specified arrhythmias and other heart rhythm disorders
- Hypertension
- Hip or vertebral fracture
- Lower-risk cardiovascular disease
- Cerebrovascular disease
- Morbid obesity
- Urinary disorders
- Psychiatric disorders other than depression

Measures of frailty/disability defined based on 1) use of durable medical equipment (DME) using Policy Group Maps maintained by Palmetto GBA under contract to CMS and 2) original reason for Medicare entitlement

- Walking aids
- Wheelchairs
- Hospital bed
- Lifts
- Oxygen
- Original reason for entitlement: disability insurance beneficiary
- Original reason for entitlement: end stage renal disease

Social risk factors

- Low AHRQ SES index
- Low physician-specialist density

Detailed Description of Hierarchical Model

The ACO MCC admission measure HGLM assumes the outcome has a known exponential family distribution and relates linearly to the covariates via a known link function, h. For the model, we assumed a negative binomial distribution with linear variance (NB-1) and a log link function (Note the NB-1 model was chosen to account for overdispersion in the data). Further, we accounted for the

clustering within ACOs by estimating an ACO-specific effect, α_i, which we assume follows a normal distribution with mean μ and variance □□2, the between-ACO variance component. The following equations define the HGLM:

$$h\big(E(Y_{ij}|\boldsymbol{Z}_{ij},\omega_i)\big) = \log\big(E(Y_{ij}|\boldsymbol{Z}_{ij},\omega_i)\big) = \alpha_i + \boldsymbol{\beta}\boldsymbol{Z}_{ij} \qquad (1)$$

where ai = m + wi ; wi ~ N (0, t 2)

$i = 1...I; j = 1...n_i$

Where Y_{ij} denotes the outcome (number of unplanned admissions per the person-years of risk exposure) for the j-th patient attributed to the i-th ACO; $\boldsymbol{Z}_{ij} = \big(Z_{1ij}, Z_{2ij,} \dots, Z_{Pij}\big)$ is a set of p patient-specific covariates derived from the data; I denotes the total number of ACOs and n_i the number of patients attributed to ACO i. The ACO-specific intercept of the i-th ACO, α_i, defined above, comprises μ, the adjusted average intercept over all ACOs in the sample, and ω_i ,the ACO-specific intercept deviation from μ. A point estimate of ω_i, greater or less than 0, determines whether the ACO's performance is worse or better compared to the adjusted average outcome.

Further, Y_{ij} can be expressed as $Y_{ij} = N_{ij} / M_{ij}$ where N_{ij} denotes the total number of admissions during the measurement period and follows the negative binomial distribution with mean, $\mu_{ij}|\boldsymbol{Z}_{ij},\omega_i$, and variance with dispersion parameter θ, $\mu_{ij}|\boldsymbol{Z}_{ij},\omega_i\ (1+\theta)$. M_{ij} denotes the person-years of risk exposure for the j-th patient attributed to the i-th ACO. Given this, we re-write Equation (1) as:

$$\log\big(E(N_{ij}|\boldsymbol{Z}_{ij},\omega_i)\big) = \alpha_i + \boldsymbol{\beta}\boldsymbol{Z}_{ij} + \log\big(M_{ij}\big) \qquad (2)$$

where $\log(M_{ij})$ becomes the offset used to correct for the time at risk.

We estimate the HGLM using Stata version 15 (StataCorp, College Station, TX) (MENBREG function).

Prevalence of each risk variable and the associated rate ratios for variables in the final risk model for the MIPS MCC measure.

Variable	**MIPS MCC Cohort n = 4,659,922: Prevalence of risk factors n (%)**	**MIPS MCC Cohort n = 4,659,922: Adjusted rate ratio (95% CI)**
Crude rate (per 100 person-years)	39.1	*
Total number of admissions	1,608,763	*
Total person time at risk (in years)	4,110,499	*
*	*	*
Demographic	*	*
Age <70 y/o	740,962 (15.9%)	*
Age 70 to <75 y/o	1,033,292 (22.2%)	1.09 (1.08, 1.10)
Age 75 to <80 y/o	966,205 (20.7%)	1.24 (1.23, 1.25)
Age 80 to <85 y/o	823,759 (17.7%)	1.44 (1.43, 1.45)
Age >=85 y/o	1,095,704 (23.5%)	1.78 (1.77, 1.80)
Nine chronic disease groups	*	*
AMI	100,719 (2.2%)	1.09 (1.08, 1.10)
ALZHEIMER'S AND RELATED DISORDERS	1,279,891 (27.5%)	1.27 (1.26, 1.27)
ATRIAL FIBRILLATION	1,167,393 (25.1%)	1.17 (1.17, 1.17)
CHRONIC KIDNEY DISEASE	2,383,858 (51.2%)	1.22 (1.21, 1.22)
COPD/ASTHMA	1,613,996 (34.6%)	1.22 (1.21, 1.22)
DEPRESSION	1,685,967 (36.2%)	1.07 (1.06, 1.07)
HEART FAILURE	1,823,667 (39.1%)	1.36 (1.36, 1.37)
STROKE/TRANSIENT ISCHEMIC ATTACK	635,160 (13.6%)	1.09 (1.08, 1.09)
DIABETES	2,717,638 (58.3%)	1.10 (1.10, 1.10)
Clinical comorbidities Defined using Condition Categories (CCs) or International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes	*	*
Dialysis status (CC 134)	89,380 (1.9%)	1.54 (1.52, 1.55)
Respiratory failure (CC 82, 83, 84)	459,865 (9.9%)	1.13 (1.12, 1.13)
Liver disease (CC 27 [remove K767], 28, 29, 30)	111,999 (2.4%)	1.23 (1.22, 1.24)
Pneumonia (CC 114, 115, 116)	714,580 (15.3%)	1.19 (1.18, 1.19)

Variable	MIPS MCC Cohort n = 4,659,922: Prevalence of risk factors n (%)	MIPS MCC Cohort n = 4,659,922: Adjusted rate ratio (95% CI)
Septicemia/shock (CC 2)	314,053 (6.7%)	1.05 (1.04, 1.06)
Marked disability/frailty (CC 21, 70, 71, 73, 157, 158, 159, 160, 161, 189, 190)	569,620 (12.2%)	1.23 (1.23, 1.24)
Hematologic/al diseases (CC 46 [remove D593], 48)	501,562 (10.8%)	1.03 (1.02, 1.03)
Advanced cancer (CC 8, 9, 10, 13)	263,183 (5.6%)	1.21 (1.20, 1.22)
Infectious and immune disorders (CC 1, 3, 4, 5 [remove A1811], 6, 47, 90)	261,668 (5.6%)	1.07 (1.06, 1.08)
Severe cognitive impairment (CC 50 [remove F05, F061, F068], 64, 65, 80)	370,777 (8.0%)	1.09 (1.09, 1.10)
Major organ transplant status (CC 132, 186)	39,216 (0.8%)	1.09 (1.08, 1.11)
Pulmonary heart disease (ICD-10-CM I2601, I2602, I2609, I270, I271, I272, I2789, I2781, I279, I280, I281, I288, I289)	197,778 (4.2%)	1.14 (1.14, 1.15)
Cardiomyopathy (ICD-10-CM I420, I421, I422, I425, I426, I427, I428, I429, I43, I514, I515)	397,841 (8.5%)	1.08 (1.08, 1.09)
Gastrointestinal disease (CC 31, 32, 33, 35, 36)	993,104 (21.3%)	1.06 (1.06, 1.07)
Iron deficiency anemia (CC 49)	2,058,339 (44.2%)	1.13 (1.13, 1.14)
Ischemic heart disease except AMI (CC 87, 88, 89, 98; add ICD-10 I511, I512)	2,415,379 (51.8%)	1.15 (1.14, 1.15)
Other lung disorders (CC 112 [remove J470, J471, J479], 118)	1,939,225 (41.6%)	1.02 (1.01, 1.02)
Vascular or circulatory disease (CC 106, 107, 108, 109 [remove I701, I722])	2,220,460 (47.7%)	1.13 (1.13, 1.14)
Other significant endocrine disorders (CC 23 [remove E748, N251, N2581])	278,126 (6.0%)	1.03 (1.03, 1.04)
Other disabilities and paralysis (CC 72, 74, 103, 104, 119)	292,693 (6.3%)	1.08 (1.08, 1.09)
Substance abuse (CC 54, 55, 56)	578,732 (12.4%)	1.21 (1.21, 1.22)
Other neurologic disorders (75, 77, 78, 79, 81, 105)	1,565,850 (33.6%)	1.09 (1.09, 1.10)
Specified arrhythmias and other heart rhythm disorders (CC 96 [remove I480, I481, I482, I4891] and 97)	1,412,343 (30.3%)	1.05 (1.05, 1.05)
Hypertension (CC 95)	4,204,973 (90.2%)	1.06 (1.05, 1.07)
Hip or vertebral fracture (CC 169, 170)	240,679 (5.2%)	1.07 (1.06, 1.08)

Variable	MIPS MCC Cohort n = 4,659,922: Prevalence of risk factors n (%)	MIPS MCC Cohort n = 4,659,922: Adjusted rate ratio (95% CI)
Lower-risk cardiovascular disease (CC 91, 92, 93)	1,260,360 (27.0%)	1.03 (1.02, 1.03)
Cerebrovascular disease (CC 102 [remove I6789])	267,201 (5.7%)	1.06 (1.05, 1.06)
Morbid obesity (ICD-10-CM E6601, Z6835, Z6836, Z6837, Z6838, Z6839, Z6841, Z6842, Z6843, Z6844, Z6845)	600,726 (12.9%)	1.04 (1.04, 1.05)
Urinary disorders (CC 142 [remove N131, N132, N1330, N1339, Q620, Q6210, Q6211, Q6212, Q622, Q6231, Q6232, Q6239] and 145 [remove N2589, N259, N261, N269, Q6102, Q612, Q613, Q614, Q615, Q618])	1,370,375 (29.4%)	1.05 (1.04, 1.05)
Psychiatric disorders other than depression (CC 57, 59, 60, 62, 63 [remove F4321])	1,332,385 (28.6%)	1.08 (1.07, 1.08)
Frailty indicators Defined using Noridian Policy Groups for DME or original reason for Medicare entitlement	*	*
Walking aids	231,405 (5.0%)	0.98 (0.98, 0.99)
Wheelchairs	193,552 (4.2%)	1.13 (1.12, 1.14)
Hospital bed	75,885 (1.6%)	1.09 (1.08, 1.10)
Lifts	17,136 (0.4%)	1.03 (1.01, 1.05)
Oxygen	383,219 (8.2%)	1.38 (1.38, 1.39)
Original Reason for entitlement: DIB (may or may not have ESRD)	685,924 (14.7%)	1.25 (1.24, 1.26)
Original Reason for entitlement: ESRD (may or may not have DIB)	19,072 (0.4%)	1.24 (1.21, 1.27)
Social risk factors	*	*
Low AHRQ SES index score (<=25th pct)	847,802 (18.2%)	1.08 (1.07, 1.08)
Low specialist density (<=25th pct)	167,684 (3.6%)	1.04 (1.03, 1.05)

*cell intentionally left blank

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.

Not applicable, 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?

The approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes." [1]

The overall goal of our risk adjustment is to ensure that the measure fairly accounts for patient mix across MIPS providers. Hence, the measure risk adjusts to account for factors that are associated with the outcome (that is, acute unplanned hospital admissions), vary across MIPS providers, and are unrelated to quality of care so that measure scores reflect differences in care quality. With input from stakeholders, we built off of work done with CMS on the original ACO MCC admission measure (ACO-38/NQF #2888) to create a conceptual model for risk adjustment and selected potential factors for adjustment. [Note that an updated version of the ACO MCC admission measure, which has been aligned with this MIPS MCC measure, has been submitted for endorsement maintenance in this same NQF cycle.] In addition to demographic and clinical variables, these factors included novel approaches to adjusting for patients with markers for medical frailty as well as adjusting for patient social risk factors. After selecting candidate factors, we finalized our list using the analysis described below. Finally, we reviewed the model used to calculate a risk-standardized acute admission rate (RSAAR) for MIPS providers.

Conceptual Model for Risk Adjustment

The MIPS MCC measure is built as an adaptation of a similar measure developed for CMS identifying acute admission rates for MCC patients in the ACO setting [2]. Building on the conceptual model developed in that measure, we defined and illustrated the potential relationships between different categories of risk factors and the outcome of hospital admissions. This MIPS conceptual model (see the figure below) guided the selection of candidate risk factors. We identified patient demographic factors and clinical variables, including comorbidities and measures of frailty and disability, which reflect the characteristics of the patients at the start of the measurement year and are independent of quality of care. The potential clinical variables included not only clinical comorbidities but also measures of disease severity and frailty/functional status.

We also considered social risk factors that may influence patients' risk of acute, unplanned admissions. There are many ways to conceptualize or categorize social risk factors. We adopted the model of the National Academies of Sciences, Engineering, and Medicine (NASEM) comprehensive, expert report of 2017, in which they categorized social risk factors into the following four domains. [3]

Socioeconomic position;

Race, ethnicity, and cultural factors;

Social relationships; and

Residential and community context.

(Note: There is a fifth domain in the NASEM report related to gender and sexual orientation; however, we have left it out because the authors noted that more research is needed to understand the relationship of these factors to outcomes and because of lack of available data.)

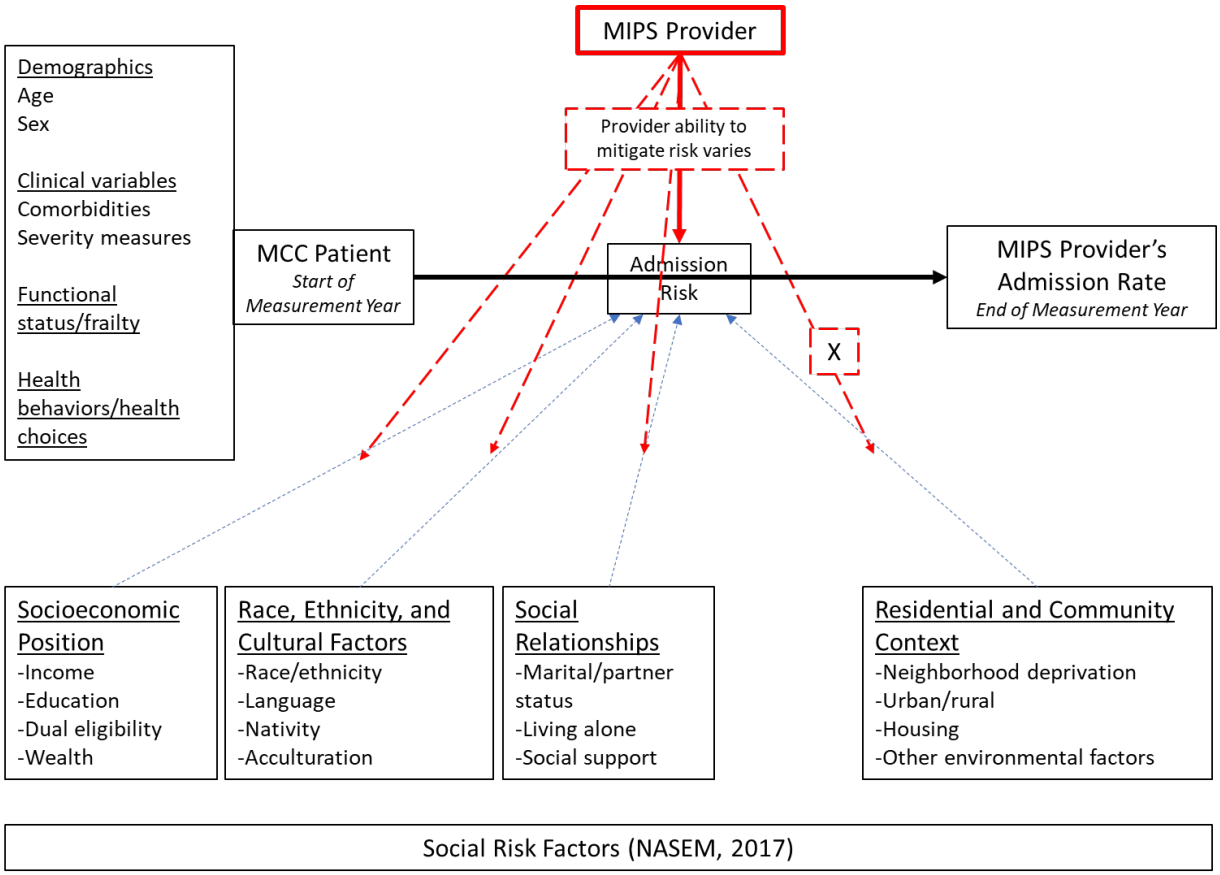


Figure 1: Conceptual model for risk adjustment

As noted in our conceptual model (Figure 1), variables in all of these domains are or are hypothesized to be associated with increased risk of admission. The domains differ, however, in the extent to which we expect an individual MIPS clinician or group of clinicians to be able to mitigate the risk conferred by such variables. These differences inform their potential use as risk adjusters, since adjusting for factors that can more easily be mitigated by higher-quality care is more likely to mask low-quality care.

MIPS providers have the least ability to mitigate the risk of admission associated with broader residential and community factors, such as neighborhood deprivation and relative lack of access to primary and specialty medical care. In contrast, however, we expect that there is more although not unlimited ability for a MIPS provider to intervene to mitigate some or all of the risk conferred by the other, individual-level domains noted above. For example, a provider can take into account a patient's education level, health literacy level, and home living situation when planning and delivering care. In addition, high-quality care may be characterized as being more racially, linguistically, and culturally sensitive and informed. While such tailored care can likely mitigate risk of admission, our TEP emphasized that providing it also requires resources, so MIPS providers may be limited in their capacity to deliver it.

Identification of Candidate Demographic and Clinical Variables

To represent demographic and clinical risk factors, the candidate variables were:

Age;

Indicator variables for each of the nine MCC cohort-qualifying conditions;

Clinical risk adjusters from the ACO MCC admission measure that we are adapting for the MIPS program;

Any additional clinical risk adjusters from the ACO diabetes admission measure not already captured by the ACO MCC admission measure, since we added diabetes as a cohort-qualifying condition; and

Measures of frailty/disability based on: a) use of selected durable medical equipment (DME) and b) original reason for Medicare entitlement.

Each of these sets of variables is described below. See Table 5 of the accompanying data dictionary for the complete list of 54 candidate demographic and clinical risk-adjustment variables and the specific Condition Category (CC) and ICD-CM codes used to define the clinical variables.

Age (1 variable with 5 levels)

In terms of demographic variables, we included age but not sex, consistent with the rationale and approach taken by CMS for other outcome measures. Studies suggest that sex-based differences in outcomes are generally driven by age and comorbidities (which we include in the risk adjustment), as well as disparities in care delivery (for example, women with diabetes and heart failure tend to receive less evidence-based treatment), and not by biological differences. [4-7] (Note: Race/ethnicity is addressed below under final risk adjustment variable selection.)

Indicator variables for the MCC cohort-qualifying conditions (9 variables)

We included indicator variables for each of the nine cohort-qualifying conditions in order to adjust for differences in risk of admission across conditions.

Clinical risk adjusters from the ACO MCC admission measure (34 candidate variables)

Final risk adjusters for the ACO MCC admission measure that we are adapting for the MIPS program include 35 clinical comorbidities (specified using CC v22).

Additional clinical risk adjusters from the ACO diabetes admission measure (3 candidate variables)

In the cohort, all chronic conditions are defined using the CCW definition with the exception of diabetes, which uses the cohort definition from the ACO diabetes admission measure developed by CORE (v2018a ACO-36). Likewise, in order to capture potential risk-adjustment variables that might be specific to diabetes, we compared the clinical risk adjusters for the ACO diabetes and the ACO MCC admission measures and added to the list of candidates any that were on the diabetes, but not MCC, list. Specifically, we added "Other organ transplants" (CC 187) to the list of candidate risk adjusters as it is a clinical comorbidity for the ACO diabetes, but not the ACO MCC measure.

In addition, the diabetes measure adjusts for disease severity using the Diabetes Complications Severity Index (DCSI). [7] Based on our team's clinical review, we noted that – with the exception of conditions related to diabetic retinopathy and precerebral arterial occlusion and transient cerebral ischemia without infarction – all of the conditions in the DCSI were already captured by the risk-adjustment variables noted above. Thus, we added these additional two clinical comorbidities to the list of candidate risk factors (see Table 5 in the accompanying data dictionary for the specific codes used to define the variables).

Measures of frailty/disability (7 candidate variables)

a) Frailty based on DME (5 candidate variables): Frail elderly patients are at increased risk of hospitalization. [8,9] The list of candidate clinical comorbidities already includes CC-defined diagnosis variables related to frailty, including "Marked disability/frailty" (CCs 21, 70, 71, 73, 157-161, 189, 190) and "Hip or vertebral fracture" (CCs 169, 170). In addition, we also considered five variables that capture past-year use of walking aids, wheelchairs, home hospital beds, lifts, and home oxygen, which can serve as claims-based proxy measures of frailty and are based on HCPCS codes for DME. [7,8] We defined these variables using the Policy Group Map currently maintained by Palmetto GBA under contract to

CMS. [10] (See Table 5 and Table 8 in the accompanying data dictionary for the codes used to define the variables).

b) Original reason for Medicare entitlement (2 candidate variables): Individuals may enroll in Medicare before the age of 65 years if they qualify for disability insurance benefits (DIB) and/or have end-stage renal disease (ESRD). Those whose original reason for Medicare entitlement was one or both conditions (as opposed to aging in) represent a particularly vulnerable, at-risk group. Thus, we additionally considered original entitlement reason as potential risk adjusters.

Identification of Candidate Social Risk Factors

Risk Adjusting for Social Risk Factors

In developing and evaluating social risk factors for potential inclusion in the model, we considered with CMS and stakeholders the conceptual model above, the program context and the pros and cons of the alternatives. As noted above, in the MIPS setting, smaller provider groups may have a limited ability to influence the broader residential and community factors that affect admission risk. Hence, we evaluated such factors for adjustment.

However, there are several potential downsides to adjusting for SRFs. Statistically, adjusting will set different expected rates of admission for different patient groups (that is, higher expected admission rates for patients with SRFs). This approach could contribute to greater acceptance of higher admission rates for people with SRFs. Adjusting could also mask quality differences associated with the risk factor. If people with MCCs and the risk factor systematically receive poorer quality care, and their admission rates are higher as a result of that worse care, adjusting for the SRF will make that worse care less visible in the measure score.

On the other hand, there are potential unintended consequences of not adjusting. In a mandatory program such as MIPS, if these factors strongly influence the outcome, not adjusting for them could result in measure scores that translate into downward Medicare payment adjustments for providers serving patients with social risk factors. If the lower scores reflected case mix rather than quality, it would not advance MIPS policy goals. Further, not adjusting might reduce resources among the providers already facing the largest resource constraints. Moreover, if providers anticipate a poor score by the measure may further reduce their Medicare payments, the measure could create an incentive to reduce access to care for vulnerable patients.

To inform further consideration of these tradeoffs, we examined the marginal effects social risk factors have after adjusting for demographic and clinical variables. If, after adjusting for demographic, comorbidity, and frailty/disability factors, social risk factors still have an independent relationship to the outcome, the balance of these concerns may tip toward adjusting for social risk factors in the context of a quality program assessing the quality of outpatient care by MIPS providers, especially for those social risk factors that providers have less of an ability to address.

For further context, we note that the MIPS program has a "complex patient" policy adjustment that is intended to address the challenges with caring for patients with social risk factors; the policy adjustment raises the score used to calculate provider payment increases/decreases by up to 5 points (out of a 100 maximum). It allocates points based on each provider's proportion of patients who are dual-eligible and their patients' average risk score (82 FR 53771 through 53776).

Development of Candidate Social Risk Factors

As with the demographic and clinical variables above, we explored available data that could be used to operationalize candidate social risk factors guided by our conceptual model. Based on our team's prior work and a focused literature review, we began with an initial list of 33 potential social risk factors,

which we narrowed to five. We used the following considerations to help guide candidate variable construction:

If an individual-level variable was not available, we considered whether a corresponding area-level variable may be a reasonable alternative.

If a construct (for example, housing) could be defined using a range of variables and there was no clear rationale for why any particular variable(s) would be associated with increased risk of hospitalization, we favored using a summary variable that captures multiple aspects of the construct to avoid a multiple-comparisons problem and spurious findings.

We focused on variables that would be expected to have a more direct impact on the outcome of interest. For example, social capital has been shown to be associated with a broad range of individual- and community-level health outcomes. [11] However, there is no clear relationship with unplanned hospital admissions; any hypothesized or observed association would be mediated by intermediate factors.

Evaluating potential data sources and variables for each of these domains yielded 5 candidate variables.

Socioeconomic position (1 candidate variable)

The focus of this domain is on individual-level measures of patients' SES. Direct measures of Medicare beneficiaries' income, wealth, and education are not available. Consistent with the NASEM model, we included the readily available and widely used dual-eligibility status variable as it is a marker of low income and assets. We defined Medicare beneficiaries as being dual-eligible if they were enrolled in Medicaid with full benefits for at least 3 months during the measurement year or the 6 months prior.

Note: Area-level measures of income, education, and assets (for example, home ownership) are available through the US Census and are discussed below under the residential and community context domain.

Race, ethnicity, and cultural factors (0 candidate variables)

We did not include any candidate variables from this domain. Except for race, data are not available for variables such as language, immigrant status, and acculturation. Lack of data notwithstanding, we do not want to adjust for factors associated with care quality. As noted in our conceptual model, we expect that there is an ability on the part of MIPS providers – at least to some extent – to intervene on the risk conferred by such variables. For example, high-quality care may be characterized as being more racially, linguistically, and culturally sensitive and informed. Moreover, consistent with guidance from our TEP and CMS, we did not consider race as a final candidate social risk factor to avoid setting different standards of care for different groups of patients.

Social relationships (0 candidate variables)

Similarly, we did not include any candidate variables from this domain. Individual-level measures of variables such as marital/partner status and social support are not available for Medicare beneficiaries. Although some area-level Census variables are available for related concepts (for example, percentage of residents never married or percentage living alone), they are not available at a more granular level (for example, 9-digit ZIP code level).

Residential and community context (4 candidate variables)

In contrast to the domains above, we conceptualized that there is less ability on the part of a MIPS provider to mitigate the risk of admission associated with broader residential and community factors. In total, we included 4 candidate variables that reflect aspects of neighborhood deprivation (AHRQ SES

Index), place of residence (rurality), and access to care (measures of PCP and physician-specialist density).

The AHRQ SES Index is a widely used variable that summarizes area-level measures of employment, income, education, and housing. In our team's previous work and the work of others, various aspects of income (for example, household income, poverty rate, income inequality) and housing (for example, value, ownership, crowding) have been examined in relation to quality measurement. Because there is no hypothesized reason specifically supporting the use of any particular neighborhood variable(s) for this measure of unplanned hospital visits, we favored the use of a composite variable that was more likely to capture relative SES across neighborhoods. Each of the index components is available at the census block level, which we then used to link to patient's residence using 9-digit ZIP code.

Consistent with the NASEM model, we also categorized beneficiaries' place of residence in terms of rurality, given its implications for timely receipt of care and concerns that individuals in more rural areas may suffer delays due to longer travel distance and time and relative lack of providers.

To more fully and directly characterize access to care, we additionally included as candidate variables two measures of provider density: 1) PCP density and 2) physician-specialist density.

Final Risk Adjustment Variable Selection

Selection of Demographic and Clinical Variables

For development and testing of the patient-level model, we randomly split the 2015 Medicare MCC Full Sample into Development and Validation samples. The Development Sample included a random 50% sample, and the Validation Sample included the remaining 50% of MCC patients not selected into the Development Sample.

Prior to variable selection, we first evaluated the prevalence and bivariate relationship between each candidate risk variable and the outcome using the 2015 Medicare MCC Full Sample. Candidate variables with a prevalence less than 0.5% or a rate ratio (RR) less than 1.3 were not considered for model selection. To select the final set of demographic and clinical variables to include in the risk-adjustment model, we performed backward variable selection on bootstrap samples. Briefly, 1,000 samples were selected with replacement from the Development Sample. For each of the 1,000 samples, a parsimonious negative binomial regression model was selected by iteratively removing non-significant candidate variables from the model using backward selection approach. All variables significant at $p<0.05$ were retained in the preliminary model. This approach yielded 1,000 models from which we then selected all variables that were retained in the model at least 90% of the time for inclusion in the measure's preliminary risk model. The 90% cut-off was selected as a more conservative inclusion criterion due to the large sample size of the measure's cohort. This method selects variables that reliably and consistently enter the model across the 1,000 bootstrap samples.

Evaluation of Social Risk Factors

We examined the marginal effects of adding social risk factors to the model, after adjusting for the selected demographic and clinical factors, using the 2015 Medicare MCC Full Sample. Given our conceptual model and the MIPS policy adjustment for dual eligibility, we took a phased approach. We first considered the added contribution of the residential and community context variables and then sought to determine the incremental effect of dual-eligibility status above and beyond all other variables in the model. We began by analyzing the 4 residential and community context variables since MIPS providers are less able to mitigate the risk of admission associated with this domain of variables, and thus accounting for these variables would contribute to greater fairness in measure score

comparisons across MIPS providers. We then examined the marginal impact of adding dual-eligibility status to the model after all other variables had been added.

For each of the continuous variables (AHRQ SES Index and the two provider density variables for PCPs and specialists), we first categorized the variable into deciles and examined the outcome rate by decile to determine whether its effect on admission rates was linear or non-linear and, if non-linear, where appropriate cut points would be. Based on this preliminary analysis, we dichotomized each variable as the lowest quartile (Q1) vs. above (Q2-Q4).

We then analyzed the four residential and community context variables and dual eligibility using negative binomial regression modeling:

For each social risk factor, we calculated its univariate rate ratio (RR), which reflects how much more the admission rate is for individuals with a given social risk factor than those without the social risk factor.

We then calculated the RR adjusted for the demographic and clinical variables.

Statistically significant social risk factors with an adjusted RR≥1.05 were included in a multivariable model that included these social risk factors and the demographic and clinical risk factors.

If the social risk factor continued to be statistically significant even after adjusting for all other variables in the multivariable model, we considered its inclusion in the preliminary model given our conceptual model of providers' ability to mitigate social risk factors and the program context.

Model Form

The measure calculates risk-standardized acute admission rates (RSAARs) for MIPS providers at the TIN level based on their MCC patients' unplanned hospital admissions during the measurement period. The RSAAR for each MIPS provider is calculated as the ratio of the number of "predicted" to the number of "expected" admissions per 100 person-years, multiplied by the national rate of admissions among all attributed Medicare FFS patients with MCCs. The measure uses a hierarchical (two-level) negative binomial model with linear variance that adjusts for demographic, clinical, and social risk factors; accounts for the clustering of patients within MIPS providers; and accommodates the varying MCC patient population size across MIPS providers.

We selected the model form based on statistical considerations. Since the outcome of number of acute hospital admissions was over dispersed (variance of the outcome exceeded the mean), we considered five alternative models: 1) negative binomial with quadratic variance or NB-2, 2) negative binomial with linear variance or NB-1, 3) zero-inflated negative binomial, 4) generalized Poisson, and 5) Poisson inverse Gaussian models. For each model form, we examined fit through use of 1) an internal calibration plot that compared observed and predicted admission rates across deciles of admission risk and 2) goodness-of-fit statistics, including the Akaike information criterion (AIC), using the Development Sample. We also tested whether the inclusion of an interaction term between age and number of cohort-qualifying chronic conditions significantly improved each model form's fit, which it did not.

Citations

1. Krumholz HM, Brindis RG, Brush JE, et al. 2006. 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 113: 456-462.
2. Centers for Medicare & Medicaid Services. #2888 Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions, Last Updated: Jan 09, 2019. 2016;

http://www.qualityforum.org/ProjectTemplateDownload.aspx?SubmissionID=2888. Accessed February 20, 2019.
3. Steinwachs DM, Stratton, K., Kwan, L. Y.,. Accounting for Social Risk Factors in Medicare Payment. Washington DC: 2017 by the National Academy of Sciences;2017.
4. Perreault L, Ma Y, Dagogo-Jack S, et al. Sex differences in diabetes risk and the effect of intensive lifestyle modification in the Diabetes Prevention Program. Diabetes Care. 2008;31(7):1416-1421.
5. Vaccarino V, Chen YT, Wang Y, Radford MJ, Krumholz HM. Sex differences in the clinical care and outcomes of congestive heart failure in the elderly. Am Heart J. 1999;138(5 Pt 1):835-842.
6. Wexler DJ, Grant RW, Meigs JB, Nathan DM, Cagliero E. Sex disparities in treatment of cardiac risk factors in patients with type 2 diabetes. Diabetes Care. 2005;28(3):514-520.
7. Young BA, Lin E, Von Korff M, et al. Diabetes complications severity index and risk of mortality, hospitalization, and healthcare utilization. Am J Manag Care. 2008;14(1):15-23.
8. Faurot KR, Jonsson Funk M, Pate V, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. Pharmacoepidemiol Drug Saf. 2015;24(1):59-66.
9. Kim DH, Schneeweiss S, Glynn RJ, Lipsitz LA, Rockwood K, Avorn J. Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index. J Gerontol A Biol Sci Med Sci. 2018;73(7):980-987.
10. Palmetto GBA. Policy Group Map Reports. 2019; https://www.dmepdac.com/palmetto/PDAC.nsf/DID/CVAIRKXWKW. Accessed February 20, 2019.
11. Kawachi I. Social capital and community effects on population and individual health. Ann N Y Acad Sci. 1999;896:120-130.

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?

In total, we began with 54 candidate demographic and clinical variables and 33 potential candidate social risk factors. Based on our variable selection criteria and statistical methods detailed above in Section 2b3.3a, we excluded a total of 7 candidate clinical variables because they had a prevalence of less than 0.5% (1 variable), had an unadjusted rate ratio of less than 1.3 (3 variables), or were not retained in at least 90% of the bootstrap results (3 variables), leaving 47 demographic and clinical variables in the final model. As detailed above in Section 2b3.3a, once these patient factors were selected, we took a phased approach to evaluating and selecting the social risk factors. After adjusting for the demographic and clinical variables as well as each other, 2 social risk factors were retained in the final model. Thus, the final risk-adjustment model included 47 demographic and clinical variables and 2 social risk factors (see Section 2b3.1.1 for the list of variables in the final model).

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.**

Consistent with our conceptual model and statistical approach to risk factor selection, we include 2 social risk factors in the risk-adjustment model: AHRQ SES Index and physician-specialist density.

As noted in 2b.4.3, during measure development (using the 2015 Medicare MCC Full Sample) we evaluated four residential and community context variables for possible inclusion in the risk-adjustment model: 1) the AHRQ SES Index, 2) rural residence, 3) PCP density, and 4) physician-specialist density. Of the four, only AHRQ SES Index and physician-specialist density met our thresholds of having a demographic and clinical variables-adjusted RR of ≥1.05 and remaining statistically significant at the 0.05 level in the multivariable model that included all of the risk adjusters. When the demographic and clinical variables, AHRQ SES Index, and physician-specialist density were included in the model, the adjusted RRs for AHRQ SES Index and physician-specialist density were 1.07 (95% CI: 1.07-1.08) and 1.04 (95% CI: 1.04-1.05), respectively; when dual eligibility was also added to the model, the RR for AHRQ SES Index was 1.06 (95% CI: 1.06-1.06), and the RR for physician-specialist density increased slightly at 1.05 (95% CI: 1.04-1.06).

Dual eligibility was strongly predictive of admissions (RR=1.42, 95% CI: 1.42-1.43) in the univariate model. However, the effect of dual-eligibility status was greatly attenuated by adjustment for the demographic and clinical characteristics (RR=1.12, 95%: 1.12-1.13), and relatively unchanged when the other two social risk factors were also in the model (RR=1.11, 95% CI: 1.10-1.11).

We found similar results for all three social risk factors when we updated the testing results for the selected risk factors in ICD-10-coded data, using the ICD-10 Testing Dataset (2018 data) (Table 2).

Table 2: Social risk factor testing results in ICD-10-coded data

Social Risk Factor	Univariate model RR (unadjusted) (95% CI)	Multivariate model RR (adjusted for demographic and clinical variables) (95% CI)	Multivariate model RR (additionally adjusted for AHRQ SES index and physician-specialist density) (95% CI)	Multivariate model RR (additionally adjusted for Medicare-Medicaid dual eligibility status) (95% CI)
Low AHRQ SES Index	1.14 (1.13, 1.14)	1.08 (1.07, 1.08)	1.08 (1.07, 1.08)	1.06 (1.06, 1.07)
Low specialist density	1.05 (1.04, 1.06)	1.05 (1.04, 1.06)	1.04 (1.03, 1.05)	1.04 (1.03, 1.05)
Medicare-Medicaid dual eligibility status	1.44 (1.43, 1.45)	1.11 (1.11, 1.12)	Not applicable	1.10 (1.10, 1.11)

Based on the conceptual model, testing results, and feedback from public comment and our national TEP; CMS has decided to adjust for the AHRQ SES Index and physician-specialist density, both of which were modestly but independently associated with the risk of hospital admission. These two community context variables primarily reflect factors that individual providers are unlikely to mitigate; adjusting for them is therefore less likely to adjust away quality differences across providers.

While dual-eligible beneficiaries are likely to have fewer available health/healthcare supports, and may also have other unmeasured SRFs (e.g., low health literacy), CMS is not adjusting the model for dual eligibility because:

- Adjusting for dual eligibility can mask disparities in care for dual-eligible beneficiaries as acknowledged by one provider association in public comment.
- Clinicians may have more ability to mitigate social risk associated with dual eligibility, especially if a dual-eligible beneficiary is living in a non-socially deprived community.
- Not adjusting for dual eligibility is aligned with the conceptual model for the measure; the model developed with the TEP emphasizes adjusting for community factors rather than individual because patients living within very under-resourced areas pose challenges that are particularly hard for clinicians to address (e.g., lack of community services, transportation, poor housing, and/or low education).
- TEP members supported including only the AHRQ SES Index and physician-specialist density social risk factors in the model given these factors and the program context.
- The marginal impact of including dual eligibility is attenuated after accounting for demographic, clinical, and frailty risk factors, as well as the AHRQ SES Index and physician-specialist density social risk factors.

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*)

We assessed adequacy of the demographic and clinical risk-adjustment model first in the Development Sample and then evaluated model performance in the Validation Sample. The measure uses the number of acute unplanned hospital admissions per person-year at risk for admission as the outcome. Because the outcome is a count of hospital admissions – rather than a binary outcome, such as whether or not a patient has been admitted – several routinely used metrics of model performance cannot be applied (for example, we cannot use a c-statistic).

We computed two summary statistics for assessing the risk-adjustment model performance: 1) goodness-of-fit statistic (deviance R-squared) using both the Development and Validation Samples, as well as the ICD-10 Testing Dataset and 2) overfitting indices (Validation Sample only). We then compared the deviance R-squared in both the Development and Validation Samples. For the overfitting indices (γ 0, γ 1), we investigated how far the estimated value of γ 0 was from 0 and the estimated value of γ 1 was 1, using the Validation Sample.

Using the Full Sample, we also compared the deviance R-squared between the risk models with and without adjustment for the social risk factors.

Deviance R-squared (model discrimination)

We calculated deviance R-squared using the deviance residual defined by Cameron [1]. The deviance R-squared evaluates how successful the fit is in explaining the variation of the data. Deviance R-squared can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of deviance is accounted for by the model. For example, a deviance R-squared value of 0.12 means that the fit explains 12% of the total deviance.

Overfitting indices (model calibration)

Overfitting refers to the phenomenon in which a model accurately describes the relationship between the predictive variables and the outcome in the development dataset but fails to provide valid

predictions in new patients. We calculated overfitting indices of γ0 and γ1. Estimated values of γ0 far from 0 and estimated values of γ1 far from 1 provide evidence of overfitting.

Model performance among patients at different risk of admission (model calibration)

In order to determine whether the model performs well across groups of patients at different risk of admission, we divided the ICD-10 Testing Dataset into quartiles of predicted admission rate. We then assessed the predicted probability of the number of admissions derived from the model compared with the observed probability of the number of admissions. The predicted probability for a group of patients is the average probability of observing 0, 1, 2, …n hospital admissions, given these patients' risk factors for admission. The observed probability of each count of admissions for a group of patients is the proportion of these patients admitted to the hospital 0, 1, 2, …n times.

Citation:

1. Cameron AC and AG Windmeijer. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. J Bus & Econ Stat, 14(2):209-220, 1996.

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*)**:**

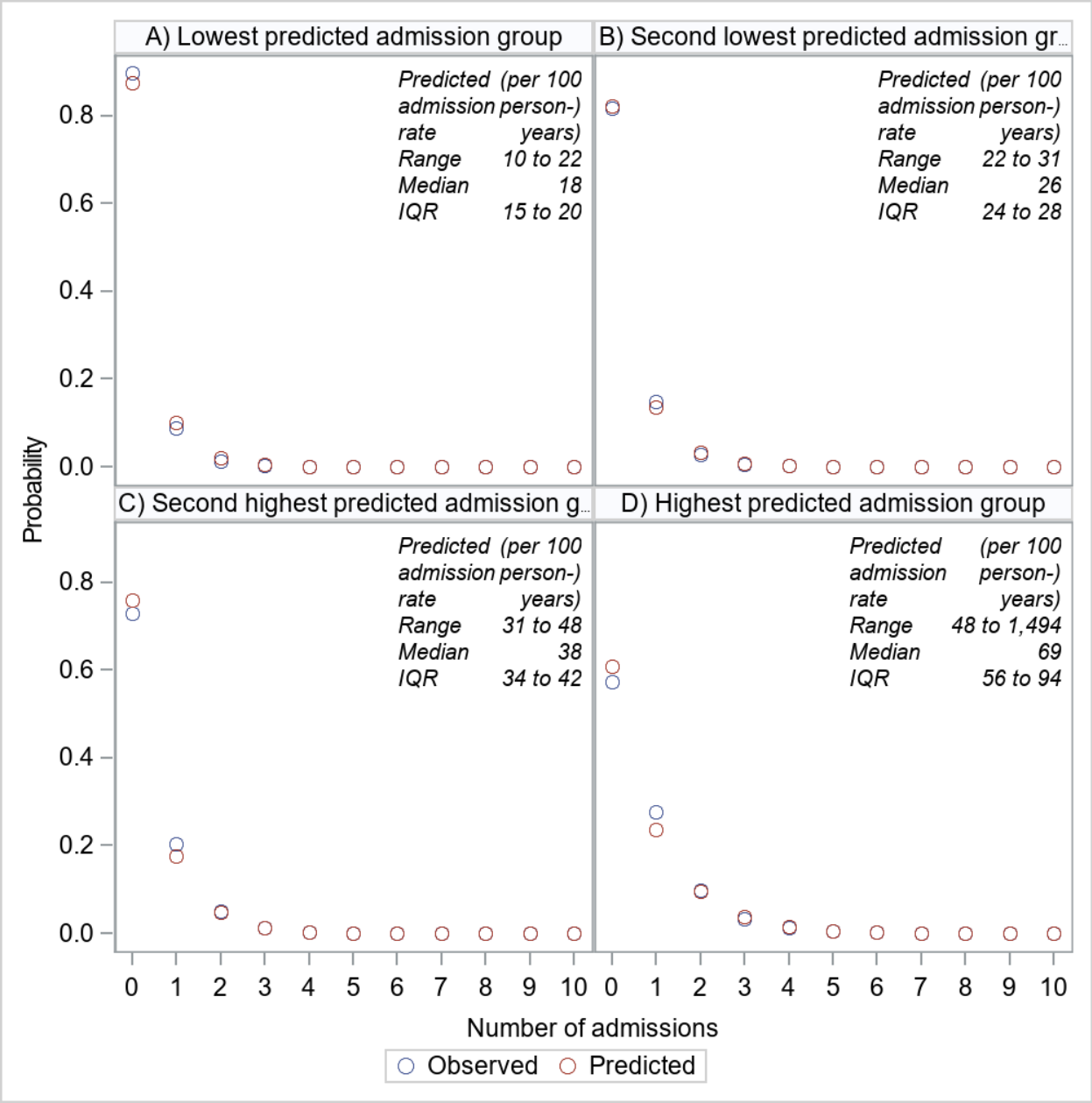
The deviance R-squared for the model with demographic and clinical risk factors was 0.105 in the Development and Validation subsamples and the 2015 Medicare MCC Full Sample, indicating that the model explains 10.5% of the variation in admission rates. The 2015 Medicare MCC Full Sample deviance R-squared was substantively similar after adding the AHRQ SES Index and physician-specialist density variables to the model (from 0.105 to 0.106); the deviance R-squared in the ICD-10 Testing Dataset (including all risk variables in the final model: demographic, clinical, and social risk factors) was similar (0.108).

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

In the Validation Sample, the over-fitting index of γ0 was close to 0 (-0.0002) and γ1 was close to 1 (0.9997), indicating good calibration of the demographic and clinical risk model.

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

A comparison of observed versus predicted probability for the number of hospital admissions among patients with multiple chronic conditions by risk quartile in the 2018 ICD-10 Testing Dataset is shown below.



The plots of observed and predicted probabilities for each number of hospital admissions (0, 1, 2, …, 10) across quartiles of risk showed that the model performs well across a broad range of risk. In the highest-risk group, we found that the observed and predicted probabilities for zero and one admission differed slightly. However, these differences were small and somewhat expected among the highest-risk group of patients.

2b3.9. Results of Risk Stratification Analysis:

Not applicable. This measure is not risk stratified.

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*)

The results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics.

Discrimination statistics: Model performance was similar in the Development and Validation Samples, with good model discrimination and fit. This held with adjustment for clinical, demographic, and social risk factors.

Calibration statistics: The calibration indices of γ0 and γ1 had values close to 0 and 1, respectively, indicating good calibration of the model with little evidence of overfitting.

Risk-decile plots: The plots, which showed that the predicted risk closely approximated the observed risk in most deciles, suggest reasonable calibration.

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*)

We found comparable risk factor frequencies and ratio ratios in the Development and Validation Samples. Please see Table 6 of the attached methodology report.

We also use this section to describe our assessment of different approaches to attribution. This information excerpted from Appendix C of the attached methodology report.

Appendix C. TIN-Level Attribution Options Considered

We considered two approaches to TIN-level attribution:

1. Patients who are assigned to an individual clinician follow that clinician to her/his TIN.
2. Patient assignment is run at the TIN level, independent of individual clinician assignment (TIN "roll-up" option).

Appendix C.1. TIN-Level Option 1

The rationale for TIN-level Option 1 wherein patients who are assigned to an individual clinician follow that clinician to her/his TIN is:

- An individual clinician should have accountability for each individual patient.
- This is aligned with most insurance policies that require patients to choose a PCP.

Key features of this approach are:

- Patients who are assigned to an individual clinician follow that clinician to her/his TIN.
- Clinicians who elect to report their quality as part of a group bring their patients with them.

There may be instances, for example, where solo provider PCP1 is assigned a patient based on having 4 visits; however, PCP2, PCP3, and PCP4 with 2 visits each who are in a TIN together collectively have more visits.

- Patients unassigned at the individual clinician level continue to be unassigned at the TIN level.

Appendix C.2. TIN-Level Option 2

The rationale for TIN-level Option 2 wherein patient assignment is run at the TIN level, independent of individual clinician assignment (TIN "roll-up" option), is:

- In a group/medical home model, clinicians assume joint responsibility for patients' care and outcome.
- Care is optimized by a team-based approach that holds clinicians jointly accountable.

Key features of this approach are:

- The patient assignment algorithm is applied to total visits among PCPs and specialists within each TIN. For example, the algorithm first identifies the TIN that billed the greatest number of PCP visits, with a minimum of 2 visits, independent of whether the patient saw a single PCP or multiple PCPs within the TIN.
- Patients unassigned at the individual clinician level may be assigned to a multi-provider TIN based on visit totals across providers in the TIN.
- This is consistent with current charge-based approach to TIN attribution under the MIPS program.

Appendix C.3. Evaluation of TIN-level Attribution Options

We evaluated the options using the criteria specified in Table C1 below. Based on TEP and CMS input and an evaluation of the selection criteria above, we selected Option 1 for attributing patients with MCCs to TINs. The approach first assigns patients to the clinician most responsible for their care (using the algorithm for individual clinician-level attribution described in Section 2.5) and then has the patient follow her/his clinician to the TIN designated by the clinician.

Table C1. Evaluation of TIN-level attribution options against selection criteria

Criterion	Option 1 (patient follows clinician)	Option 2 (TIN roll-up)
Transparency/ease of understanding	**More** When patients follow clinicians to their TINs, it is clear to clinicians and TINs where patients will be assigned at the TIN level More consistent with program structure providing clinician choice of reporting at NPI/TIN or TIN level	**Less** Multi-provider TINs will be less able to anticipate which patients will be attributed to them based on the provider composition of the TIN
Ease of implementation	**More** Need to run only 1 algorithm for clinician-level and TIN-level attribution	**Less** Need to run 2 separate algorithms for clinician-level and TIN-level attribution
Aligned with the current charge-based approach to TIN attribution under the MIPS program	**No** However, a tailored approach specific to the measure is important to ensure that patients and their outcomes are correctly attributed to providers for whom the specific measured outcome is a quality signal	**Yes**

_____

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 a TIN-level risk-standardized acute admission rate (RSAAR). We characterize the degree of variability by reporting the distribution of RSAARs across TINs and by providing the median rate ratio (MRR) [1]. The median rate ratio represents the median increase in rate of acute unplanned admission if a single patient was attributed to a higher risk TIN compared to a lower risk TIN. It is calculated by taking all possible combinations of providers, always comparing the higher-risk provider to the lower-risk provider. The MRR is interpreted as a traditional rate ratio would be.

Citations

1. Austin, PC, Stryhn, H, Leckie, G, Merlo, J. Measures of clustering and heterogeneity in multilevel Poisson regression analyses of rates/count data. Statistics in Medicine. 2018; 37: 572– 589. https://doi.org/10.1002/sim.7532

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*)

Distribution of RSAAR

The distribution of the measure score for clinician groups (TINs) with more than 15 providers and at least 18 patients with MCCs is shown below. The risk-standardized measure scores estimated using Medicare FFS data (calendar year 2018, using the ICD-10 Testing Dataset) had a median value of 40.4 admissions per 100 person-years. The measure score distribution is shown below as percentiles, and in a histogram.

Providers included: Providers with >=18 MCC patients and at clinician group size >15

N=4044

Mean (sd): 41.2 (7.7)

Min: 20.4

1st percentile: 26.7

5th percentile: 30.6

10th percentile: 32.6

25th percentile: 36.0

Median: 40.4

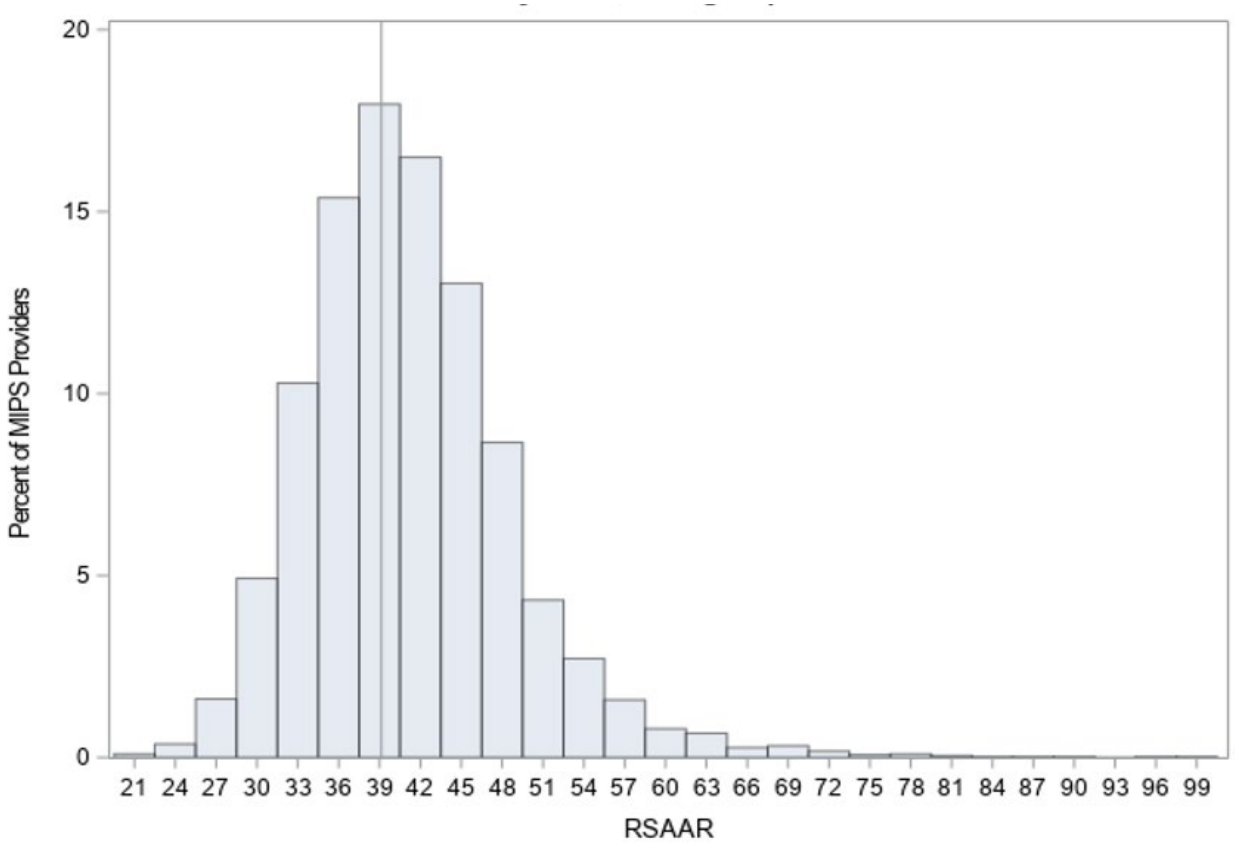
75th percentile: 45.2

90th percentile: 50.2

95th percentile: 54.7

99th percentile: 65.6

Maximum: 98.7



Median Rate Ratio (MRR) = 1.27.

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?*)

As the results above show, across the 4,044 clinician groups who had at least one MCC patient, RSAAR measure scores, including adjustment for the social risk factors of AHRQ SES Index, and physician-specialist density, ranged from 20.4 to 98.7 per 100 person-years, with a median of 40.4 and an IQR of 36.0 to 45.2. This indicates that after adjustment half of Medicare patients with multiple chronic conditions had between 36 and 45 acute care visits per 100-person years.

Furthermore, the 10th and 90th percentiles, representing the best and worst performers, had a visit rate of 32.6 and 50.2 respectively, represent meaningful deviations from the median: clinician groups in the 10th percentile had 19% fewer admissions per 100-person years compared with the median and clinician groups in the 90th percentile had 24% more admissions per 100-person years compared with the median. In addition, the best performing 5% of clinician groups had 24% fewer admissions than the median, whereas the worst-performing 5% of clinician groups (those in the 95th percentile) had 35% more admissions compared with the median.

The median rate ratios (MRRs) suggest meaningful increases in the rate of acute unplanned admissions if a single patient was attributed to a higher-risk clinician group compared to a lower-risk clinician group. At the clinician-group level, the MRR value of 1.27 indicates that a patient has a 27% higher admission rate if the patient was attributed to a higher risk clinician group compared to a lower risk clinician group indicating that the impact of quality on the outcome rate is substantial.

Overall, our results suggest that there is substantial need to both reduce the number of admissions for this patient population, as well as decrease the variation in admissions across providers, and that improvement goals are achievable.

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*)

Items 2b5.1 – 2b5.3 skipped, as this measure has only one set of specifications.

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*)

Items 2b5.1 – 2b5.3 skipped, as this measure has only one set of specifications.

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*)

Items 2b5.1 – 2b5.3 skipped, as this measure has only one set of specifications.

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*)

Demographic and clinical variables were ascertained using claims data and, as such, were known for all patients in the study cohort. Missing data for the area-level social risk factors (i.e., AHRQ SES Index and physician-specialist density) were investigated by checking whether or not the patient's ZIP code was available and could be matched to the area-level values. We also examined shifts in the quintile of TIN-level RSAARs after patients with missing data were excluded.

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*)

Of the final 4,937,865 MCC sample size in the MCC Full Sample Dataset, 21,693 (0.44%) patients had missing data for one or both social risk factors included in the risk model. These patients had P.O. boxes or business ZIP codes or lived in US territories for which data were unavailable. Across all 64,086 TINs, TINs in the 90th and 99th percentiles had 1 and 6 patients with missing data, respectively, with a maximum of 284 patients.

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*)

Excluding 0.44% of patients with missing social risk factor data did not affect the quintile of RSAARs for the majority of the providers. A shift in the quintile of RSAARs occurred only for providers with a very small patient volume and for which the measure is not meant to be reported.

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 a combination of electronic sources

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; all data elements are from electronic sources.

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 data and, as such, imposes no data collection burden to measure entities.

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; there are no fees, licensing, or other requirements to use any aspect of this measure as specified.

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)
Payment Program Not in use	*

*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

N/A; the measure is not yet in use.

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; the measure is not currently publicly reported or used in an accountability application. CMS may propose this measure for use under the Merit-based Incentive Payment System.

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; the measure is not currently publicly reported or used in an accountability application. CMS may propose this measure for use under the Merit-based Incentive Payment System.

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.

During development of the measure, CORE and CMS provided performance results, data, and/or assistance with interpretation in several ways.

1) CORE recruited and met with a national Technical Expert Panel (TEP) throughout measure development. TEP members and commenters included representatives of the measured entities and patients covered by

the measure to ensure the measure is as meaningful as possible to all stakeholders. CORE provided performance results and data to TEP members periodically for their review and input. CORE reviewed, considered, and responded to all TEP input.

2) CORE hosted a public comment after reviewing the measure with the TEP. We notified CMS listservs, CORE's stakeholders and stakeholder organization listservs including:

? Business and consumer advocacy organizations.

? Condition-related registries.

? Electronic Health Record vendors.

? Healthcare quality-focused organizations.

? Insurance and purchaser organizations.

? National professional associations and clinician societies.

? Patient advocacy groups and patient safety organizations.

? Quality improvement and measurement organizations.

? Research organizations.

? State medical societies.

? Topic knowledge-related organizations.

? The project's national Technical Expert Panel (TEP).

? TEPs and Clinician Committees for related MIPS cost or quality measures under development not covered by this project.

CORE solicited public comments on the measure, and we took all comments into consideration, addressing them individually. Therefore, performance results and data were provided to members of the TEP and then made public through public comment.

3) CORE presented the measure to clinicians and practice managers in the voluntary Clinician Champions Program to elicit feedback.

4) CORE presented the measure at national conferences (CMS Quality Conference, Academy Health).

5) CMS included the measure in pre-rulemaking (MAP) and rulemaking (proposed rule) processes. CMS added the measures to the 2019 Measures Under Consideration list for NQF Measures Application Partnership (MAP) review and included it in the Calendar Year 2020 Quality Payment Program proposed rule for stakeholder comment. The NQF MAP reviewed the measure in December 2019 as part of the 2019-2020 pre-rulemaking cycle and included it in their public comment processes as well. CORE and CMS reviewed all comments received on the measure and addressed them through the MAP review and CMS regulatory processes, respectively.

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.

CORE met with the TEP periodically and hosted a public comment period during measure development. CORE provided data and results to the TEP and obtained TEP input during four teleconference meetings throughout the measure development process and solicited TEP input via email. CORE presented the measure to the voluntary Clinician Champions and at conferences (CMS Quality Conference, Academy Health) during development, and supported the measure during the MAP's review in fall/winter 2019.

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.

The measure has not yet been implemented. During measure development, feedback was obtained as described above in 4a2.1.1.

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

The measure has not yet been implemented. Feedback during measure development included recommendations from the TEP with regards to the criteria for inclusion/exclusion of the cohort, outcome definition, risk adjustment, and attribution.

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

The measure is not currently in use. Feedback from the MAP included support for the measure concept but the MAP recommended specific changes to measure specifications: (1) measure should be specified for clinician groups only; (2) minimum reliability should be set at 0.7; (3) patient attestation should be tested by the developer as it becomes available and override claims-based attribution algorithm.

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.

In response to MAP feedback described above, the measure was defined for clinician groups and testing results were presented for clinician groups including at least 15 providers. Various cut-points for measure reliability were considered and tested, including a minimum reliability of 0.4 (at least 18 patients with MCC attributed per TIN). The reliability of 0.4 was selected to align with reliability for other MIPS measures and to optimize applicability of the measure to larger proportion of patients with MCCs, provider groups, and to optimize capture of the outcome. Patient attestation has not yet been tested.

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.

This is a new measure and there is no information available on performance improvement. This measure is not currently used in a program, but the primary goal of the measure is to provide information necessary to implement focused quality improvement efforts. Providers could use the measure information to implement practice improvements, such as those outlined in the Evidence attachment. Practice features that are associated with successfully reducing hospitalization include 1) supplementing patient telephone calls with in-person meetings; 2) occasionally meeting in person with providers; 3) acting as a communication hub for providers; 4) providing patients with evidence-based education; 5) providing strong medication management; and 6) providing comprehensive and timely transitional care after hospitalizations.

Once the measure is implemented, we plan to examine trends in improvements by comparing RSAARs over time.

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.

N/A; the measure is not yet in use.

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

N/A; the measure is not yet in use.

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)

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

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.

The measure specifications are harmonized to the fullest extent possible. The only differences are for the CMS programs and measurement levels for which they are intended: for example, the MIPS measure is attributed and scored for clinician groups under MIPS, and the ACO MCC admission measure is attributed and scored for Medicare ACOs.

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; there are no competing measures.

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.

Attachment **Attachment:** MIPSMCCMethodologyReport_v1.0.pdf

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, 203-764-5700-

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 CORE measure development team met regularly and is comprised of experts in epidemiology, internal medicine, quality outcomes measurement, and measure development. CORE convened surgical and statistical consultants with expertise relevant to outpatient surgery and quality measurement to provide input on key methodological decisions.

CORE Measure Development Team:

Mayur M. Desai, PhD, MPH – Project Lead

Kasia J. Lipska, MD, MHS – Project Lead

Faseeha K. Altaf, MPH – Project Manager

Demetri Goutos, MBA – Research Associate/Project Coordinator

Craig S. Parzynski, MS – Supervising Analyst

Andrea G. Barbo, MS – Lead Analyst

Zhenqiu Lin, PhD – Analytic Director

Jeph Herrin, PhD+ – Statistical Consultant

Megan LoDolce, MA – Contract Manager

Elizabeth E. Drye, MD, SM* – Project Director

+Flying Buttress Associates

*Yale School of Medicine

CORE convened a TEP comprised of 20 members, including clinicians, patients, and experts in quality improvement to provide input on key methodological decisions.

TEP members:

1. Mary Barton, MD, MPP; Vice President, Performance Measurement; National Committee for Quality Assurance; Washington, D.C.
2. Larry Becker, BS; Director, Strategic Partnerships, Alliances and Analytics (Retired); Xerox; Rochester, NY
3. Jacob Berman, MD, MPH; Medical Director; General Internal Medicine Center, University of Washington; Seattle, WA
4. Jane Brock, MD, MSPH; Clinical Director; Quality Innovation Network – Quality Improvement Organization National Coordinating Center, Telligen; Greenwood Village, CO
5. Brenda Cook, MSN, RN, NEA-BC; Nursing Director; Southcentral Foundation; Anchorage, AK
6. Namirah Jamshed, MBBS; Associate Professor, Division of Geriatric Medicine; University of Texas Southwestern Medical Center; Dallas, TX
7. Lorie Joseph; Patient
8. David Kraus, MD; Advanced Heart Failure and Cardiac Transplant Specialist; Stern Cardiovascular Center; Memphis, TN
9. Rozalina McCoy, MD, MS; Assistant Professor of Medicine; Mayo Clinic; Rochester, MN
10. J. Michael McWilliams, MD, PHD; Associate Professor, Health Care Policy; Harvard Medical School; Cambridge, MA
11. Amy Mullins, MD, CPE, FAAFP; Medical Director, Quality Improvement; American Academy of Family Physicians; Leawood, KS
12. Diane Padden, PhD, CRNP, FAANP; Vice President, Professional Practice & Partnerships; American Association of Nurse Practitioners; Austin, TX
13. Robert Roca, MD, MPH, MBA; Vice President/Medical Director; Sheppard Pratt Health System/American Psychiatric Association; Baltimore, MD
14. Jason Sico, MD, MHS, FAHA, FACP; Assistant Professor of Neurology and Internal Medicine; Yale School of Medicine; New Haven, CT
15. Mary Smith, DNP, FNP-BC, ONP-C, RNFA; Nurse Practitioner; Starkville Orthopedic Clinic; Starkville, MS
16. Barbara Spivak, MD; President; Mount Auburn Cambridge Independent Practice Association; Brighton, MA
17. Jennefer Watson, Patient Caregiver; Jacksonville, FL

    Daniel Weiner, MD, MS; Associate Professor of Medicine; Tufts University School of Medicine; Boston, MA
18. Roger Wells, PA-C; Family Practice and Emergency Medicine Physician Assistant; Howard County Medical Center; St. Paul, NE
19. Stephanie Wolf-Rosenblum, MD, MMM, FACP, FCCP; Physician Administrator and Vice President of Development and External Affairs; Southern New Hampshire Health System; Nashua, NH

    Patient; Participation was confidential

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

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

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

Ad.6 Copyright statement: N/A

Ad.7 Disclaimers: N/A

Ad.8 Additional Information/Comments: N/A