NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3495

Measure Title: Hospital-Wide 30-Day, All-Cause, Unplanned Readmission Rate (HWR) for the Merit-Based Incentive Payment System (MIPS) Eligible Clinician Groups

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

Brief Description of Measure: This measure is a re-specified version of the hospital-level measure, "Hospital-Wide All-Cause, Unplanned Readmission Measure" (NQF #1789), which was developed for patients who are 65 years or older, are enrolled in Fee-for-Service (FFS) Medicare and are hospitalized in non-federal hospitals.

This re-specified measure attributes hospital-wide index admissions to up to three participating MIPS Eligible Clinician Groups ("providers"), rather than to hospitals. It assesses each provider's rate of 30-day readmission, which is defined as unplanned, all-cause readmission within 30 days of hospital discharge for any eligible condition.

The measure reports a single summary risk adjusted readmission rate (RARR), derived from the volume-weighted results of five different models, one for each of the following specialty cohorts based on groups of discharge condition categories or procedure categories: surgery/gynecology; general medicine; cardiorespiratory; cardiovascular; and neurology, each of which will be described in greater detail below.

Developer Rationale: Hospital readmission, for any reason, is disruptive to patients and caregivers, costly to the healthcare system, and puts patients at additional risk of hospital-acquired infections and complications. Readmissions are also a major source of patient and family stress and may contribute substantially to loss of functional ability, particularly in older patients. Some readmissions are unavoidable and result from inevitable progression of disease or worsening of chronic conditions. However, readmissions may also result from poor quality of care or inadequate transitional or post-discharge care.

Transitional care includes effective discharge planning, transfer of information at the time of discharge, patient assessment and education, and coordination of care and monitoring in the post-discharge period. Numerous studies have found an association between quality of inpatient or transitional care and early (typically 30-day) readmission rates for a wide range of conditions.1-8 Randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates: quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; pre-discharge assessment; and coordination of care after discharge.9-17 Successful randomized trials have reduced 30-day readmission rates by 20-40%.18 Widespread application of these clinical trial interventions to general practice has also been encouraging. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions by applying lessons learned from clinical trials. Several have been notably successful in reducing readmissions within 30 days.19 Many of these study

interventions involved enhanced clinician involvement and indicate a key role for clinicians in reducing readmissions.9-17

Despite these demonstrated successful interventions, the overall national readmission rate remains high, with a within 30-day readmission following over 15% of discharges. Moreover, we show below that RARRs range from 7% to 25% for eligible clinician groups for 2015-16. Both the high baseline rate and the variability across providers speak to the need for a quality measure to prompt greater care improvement. Given that studies have shown readmissions within 30 days to be related to quality of care, that interventions, including those utilizing clinicians, have been able to reduce 30-day readmission rates for a variety of specific conditions, and that high and variable clinician-level readmission rates indicate opportunity for improvement, we sought to develop measure of all-cause, all-condition 30-day unplanned readmission at the clinician-group level.

References:

1. Frankl SE, Breeling JL, L. G. Preventability of emergent hospital readmission American Journal of Medicine. Jun 1991;90(6):667-674.

2. Corrigan J, Martin J. Identification of factors associated with hospital readmission and development of a predictive model. Health Services Research. Apr 1992;27(1):81-101.

3. Oddone E, Weinberger M, Horner M, et al. Classifying general medicine readmissions. Are they preventable? Veterans Affairs Cooperative Studies in Health Services Group on Primary Care and Hospital Readmissions. Journal of General Internal Medicine. Oct 1996;11(10):597-607.

4. Ashton C, Del Junco DJ, Souchek J, Wray N, Mansyur C. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care. Oct 1997;35(10):1044-1059.

5. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine. Apr 24, 2000;160(8):1074-1081.

6. Courtney EDJ, Ankrett S, McCollum PT. 28-Day emergency surgical re-admission rates asa clinical indicator of performance. Annals of the Royal College of Surgeons of England. Mar 2003;85(2):75-78.

7. Halfon P, Eggli Y, Pr, et al. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical Care. Nov 2006;44(11):972-981.

8. Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA. May 5, 2010;303(17):1716-1722.

9. Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M. Comprehensive discharge planning for the hospitalized elderly. A randomized clinical trial. Ann Intern Med. Jun 15 1994;120(12):999-1006.

10. Naylor MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA. 1999;281(7):613-620.

11. Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. Journal of the American College of Cardiology. Jan 2, 2002;39(1):83-89.

12. van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. Journal of General Internal Medicine. Mar 2002;17(3):186-192.

13.Conley RR, Kelly DL, Love RC, McMahon RP. Rehospitalization risk with second-generation and depot antipsychotics. Annals of Clinical Psychiatry. Mar2003;15(1):23-31.

14.Coleman EA, Smith JD, Frank JC, Min S-J, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: The Care Transitions Intervention. Journal of the American Geriatrics Society. Nov 2004;52(11):1817-1825.

15. Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. JAMA. Mar 17, 2004;291(11):1358-1367.

16. Jovicic A, Holroyd-Leduc JM, Straus SE. Effects of self-management intervention on health outcomes of patients with heart failure: a systematic review of randomized controlled trials. BMC Cardiovasc Disord. 2006; 6:43.

17. Garasen H, Windspoll R, Johnsen R. Intermediate care at a community hospital as an alternative to prolonged general hospital care for elderly patients: a randomised controlled trial. BMC Public Health. 2007; 7:69.

18. Leppin AL, Gionfriddo MR, Kessler M, etal. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Intern Med. 2014;174(7):1095-1107.

19. CFMC. CFfMC. Care Transitions QIOSC. 2010; http://www.cfmc.org/caretransitions/.

Numerator Statement: The outcome for this measure is readmission within 30-days of a hospital discharge. We define readmission as an inpatient admission for any cause, except for certain planned readmissions, within 30 days from the date of discharge from an eligible index admission.

Additional details are provided in S.5 Numerator Details

Denominator Statement: The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from any non-federal, acute care inpatient U.S. hospitals (including territories) with Medicare Part A enrollment for the 12 months prior to admission and Part A enrollment for the 30 days after discharge. These are called 'index admissions'.

Outcome attribution:

There are three eligible clinician groups for attribution: 1) the Primary Inpatient Care Provider, 2) the Discharge Clinician and 3) the Outpatient Primary Care Physician.

1) Primary Inpatient Care Provider: All patient-facing claims for the patient filed during the stay are identified and totaled by clinicians identified on each claim; the admission is attributed to the clinician with the greatest charges billed. The cost of charges billed (as opposed to number of charges) better reflects the appropriate clinician, especially for the surgical specialty cohort. The identified primary inpatient care provider may also be the discharge clinician.

2) Discharge Clinician: Identified by Current Procedural Terminology [CPT®] code 99238 or 99239 within the last three days of admission OR CPTs 99231, 99232, 99233 billed on the last day of admission. If none of these codes found, a Discharge Clinician is not assigned.

3) Outpatient Primary Care Physician: The clinician who provides the greatest number of claims for primary care services during the 12 months prior to the hospital admission date.

Eligible clinician groups are defined by grouping eligible clinicians who use the same Taxpayer Identification Number (TIN). Index admissions are attributed to a clinician group by each of these rules. Though an admission may be attributed to three distinct eligible clinician groups, it will often be the case that two or even all three of the above listed roles for a given patient are filled by clinicians assigned to the same clinician group. In the case of multiple assignments of an admission to the same eligible clinician group, each admission is included only once when measuring the eligible clinician group.

Importantly, this implies that while there are three different rules for attribution, these are not distinguished when measuring clinician group performance. While a clinician group can have admissions attributed to them in multiple capacities – for instance, a clinician from the same group may be both a Discharge Clinician for some patients and a Primary Inpatient Care Provider for others – all attributed admissions are used to construct a single score for that eligible clinician group. Thus, while we report some results by attribution role, we report measure scores only for "unique eligible clinician groups".

Additional details are provided in S.7 Denominator Details.

Denominator Exclusions: From the cohort, we exclude admissions if:

1. The patient is discharged against medical advice (AMA)
2. The patient is discharged from a PPS-exempt cancer hospital
3. The patient is admitted primarily for the medical treatment of cancer
4. The patient is admitted primarily for the treatment of psychiatric disease
5. The patient is admitted primarily for "rehabilitation care; fitting of prostheses and adjustment devices" (CCS 254)
6. Admissions without 30 Days of Post-Discharge Enrollment are excluded
7. Admissions cannot be identified in IDR database
8. The admission cannot be attributed to an eligible clinician.

Further exclusion details can be found in S.9 Denominator Exclusion Details

Measure Type: Outcome

Data Source: Claims, Other

Level of Analysis: Clinician : Group/Practice

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

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

- This is a re-specified version of the hospital-level measure, "Hospital-Wide All-Cause, Unplanned Readmission Measure" (NQF #1789). #1789 was developed for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and are hospitalized in non-federal hospitals. This specified measure attributes admissions up to three participating MIPS eligible clinicians.
- The developer also provides a logic model demonstrating clinician group-level and facility-level interventions that can be undertaken to reduce the risk of unplanned hospital visits.
- These clinician group-level factors include: ensuring appropriate discharge, medication reconciliation, reducing infection risk, and ensuring proper outpatient follow-up.
- The developer provided a citation noting that strategies used by hospitals could be adopted by clinician groups to lower readmission rates. The developer notes that medication reconciliation, discharge instructions, and outpatient follow up are examples of interventions clinicians could undertake.
- The developer provided new information on attribution to address questions raised in the last review. The measure holds three groups accountable (the clinician groups of the Primary Inpatient Clinician, Discharge Clinician, and Primary Outpatient Clinician) as a way to incentivize collaboration of care across settings.

The developer notes that each of these groups holds different responsibilities for ensuring the patient's needs are addressed and provides examples of ways each group can provide care, improve care coordination, and work on care transitions to reduce readmission rates. This revised text is in red font in the evidence section of the submission.

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

Box 1: The measure assesses a healthcare outcome →Box 2: The developer has provided empirical data that there is a relationship between the measured outcome and at least one healthcare outcome → Pass

The highest possible rating is 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.

- The developer notes the distribution of risk adjusted readmission ratios (RARRs) for for clinician groups from 13.1 in the first decile to 18.0 in the tenth decile.

Disparities

- The developer examined potential disparities affecting patients who are dually eligible for Medicare and Medicaid. The developer found a slightly higher median observed hospital visit rate of 15.6% for eligible clinicians with a high proportion (=25.8%) dual eligible patients, compared to a median observed hospital visits rate of 15.2% for eligible clinicians with a low proportion of dual eligible patients (=5.4%). However, the differences in median rates of readmissions appear minimal.

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**

RATIONALE:

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

Evidence

- Not aware of any new studies
- The developer has provided adequate evidence to support the need for a measure to monitor this important aspect of healthcare. Unplanned readmission is both complex and multi-factorial. This reviewer agrees with the move away from facility towards providers. It is not entirely clear to me however whether the three accountable groups proposed have sufficient involvement in the group-level activities of medication reconciliation, discharge instructions and outpatient follow up. Aside from medication reconciliation, the remaining two are "ordered" but not implemented or tracked by an accountable group member. This work falls to non-physicians; principally clinical nurses and case managers. This leaves this reviewer with concerns about fair attribution.

- Re: whether there is at least one thing that the provider can do to achieve a change in the measure results, there are certainly RCTs that have demonstrated an improvement in readmissions though these results are not consistent across trials. There was also a recent NEJM paper re: the very intensive intervention (Camden study) which did not demonstrate a benefit. However, for the purposes of this review, the measure receives a pass.
- How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? Data reasonably and directly supports the outcome How does the structure, process, or outcome relate to desired outcomes? reasonably well 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? no
- This re-specified measure attributes hospital-wide index admissions to up to three participating MIPS Eligible Clinician Groups ("providers"), rather than to hospitals. It assesses each provider's rate of 30-day readmission, which is defined as unplanned, all-cause readmission within 30 days of hospital discharge for any eligible condition. The developer provided a logic model that demonstrated clinician group level and facility level interventions that could be undertaken to reduce unplanned hospital visits.
- Data are adequate
- The measure is an outcome measure directly supported by multiple RCTs of interventions to decrease readmission rates. The measure is referred to as an improvement measure multiple times by the measure developers. My one concern is that this clinician group level measure is intended for use in the MIPS Program which is a VBP program. For example, in VBP the physician groups using hospitals with an approach similar to Project Red at discharge would be at a distinct advantage to perform well on this measure.
- Yes, the evidence does relate to the outcome being measured
- There is a good body of evidence on the effectiveness of interventions, particularly related to self-care, to reduce readmissions. May be useful to note the recent negative navigator study in NEJM (PMID: 31914242)
- Evidence relates, however clarity is needed regarding the provider and admit definitions.
- I think this measure has some evidence built around it, but I am struggling with the consistency and accuracy of the emprical data required for this measure
- The evidence applies directly. There are additional references that should be incorporated.

Performance Gap

- Yes performance gap provided less than optimal performance noted
- The provider adequately demonstrated a gap in care that warrants national performance measure monitoring. I am not convinced that adequate attention has been paid to Social Determinants of Health. I would like greater clarity that transportation, access to healthcare, including availability of providers and community health programs, food insecurity, education and economic challenges related to prescription costs have been considered in risk-adjustment.
- There is some variability in performance. Re: disparities, again, the developers are using MCD eligibility as a marker. We have and could continue to discuss that in terms of appropriateness.
- Was current performance data on the measure provided? yes How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? variable re-admission rates for clinicians and groups per available clams data Disparities: Was data on the measure by population subgroups provided? yes How does it demonstrate disparities in the care? not very clear on validity of this data
- The developer noted that the distribution of risk adjusted ratios for the clinician group had a range.
- Data are adequate
- Current performance on the measure is provided and supports a gap both in performance variability and less than optimal performance on the measure. A national quality improvement

measure for all cause readmission rates is warranted. Data is provided on disparities by calculating the proportion of dual eligibles and non-dual eligibles in clinician groups and by looking at eligible clinician groups based on low and high proportion of AHRQ SES index score breakpoint. Subpopulation differences are shown on both of these factors. There is no testing for other common social risk factors.

- Current data were provided, and they do describe a performance gap. This does warrant a national performance measure because of the wide range of patients it impacts
- While the RARR does vary, I would like more information about the number of readmissions that would distinguish clinician groups. What is the difference in outcome event rate at the clinician/ clinician group level? I see on page 48 the case volume is reported in the 25-200 range-- does that refer to the admissions per condition per clinician group per year?
- centered around dual eligible only.
- Yes
- Current performance data on the measure was provided and demonstrates a rationale for a national performance measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

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

- **The SMP review below refers to the previous version of the measure, evaluated in the Spring 2019 cycle. That version included an individual clinician level of analysis in addition to the clinician groups level of analysis. The individual clinician has been removed from this (Fall 2019) version. The sub-bullets immediately below are new data provided on the clinician groups LOA, reviewed by NQF staff (not the SMP). The new information is included in the testing section in red text.**
    - Developer provided new signal to noise reliability testing for each of the five specialty cohorts that are combined. The mean results range from 0.45-0.65, for a moderate to substantilal agreement.

- The developer also provided additional information on the TEP's comments on the face validity of the measure.
- The developer provided additional information on the risk adjustment testing. The developer notes that "relationships between patient social risk factors and readmissions are multifaceted. Causal pathways include patient health upon admission, social risk factors outside of the hospital, care quality of the hospital, and differential care within a hospital."

Evaluators: Christie Teigland, Karen Joynt Maddon, Susan White, Ron Walters, Jen Perloff, Jack Needleman

Evaluation of Reliability and Validity (and composite construction, if applicable):

- Summary of Methods Panel Review Process
  - This measure was evaluated by the Scientific Methods Panel but was not discussed on their call as the group came to consensus during their preliminary review. A summary of the measure and the Panel's review is provided below.
  - The SMP subgroup members who reviewed the measure agreed that the reliability testing methodology was appropriate.
  - The SMP subgroup members raised concerns that social risk factors were excluded from the risk model given the effect size and the potential for negative consequences on access to care if this measure is not adequately risk adjusted. Panel members suggested the developer examine other clinical variables that could underlie disparities such as frailty or functional status.
  - Overall, SMP subgroup members indicated the approach to validity testing was appropriate. One subgroup member questioned why the developer only provided a comparison to star ratings as an ANOVA test would have been feasible.
  - SMP subgroup members found the results of validity testing to be reasonable and the results demonstrated moderate validity. Members noted the TEP found the measure to be valid but results were somewhat moderate.

Standing Committee Action Item(s):

- The Standing Committee will discuss and vote on reliability
- The Standing Committee will discuss and vote on validity.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Evaluation: Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3495

Measure Title: Hospital-Wide, 30-Day, All-Cause, Unplanned Readmission (HWR) Rate for the Merit-Based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups

Type of measure:

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

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

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**

☒**Enrollment Data** ☐**Other**

Level of Analysis:

☒ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SUMMARY

- Reliability testing was performed for the measure score via signal-to-noise analysis
- The developer used the formula presented by Adams et al, to calculate reliability. The mean signal-to-noise reliability scores of the five cohorts ranged from 0.45 to 0.65 for clinician groups. Signal-to-noise was calculated with one year of data and providers caring for at least 25 patients in the cohorts.
- The developer provided updated reliability testing information at various case volumes. At n=25-100, the signal-to-noise reliability scores of the five cohorts ranges from 0.31-0.52.

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**
    - **Reviewer 1:**Basic model for outcome, risk adjustment, inclusions and excllusions is model used for NQF endorsed 30 day readmission for hospitals. This measure applies at the physician and physician group level. Key issues are: attribution, weighting up all cases per clinician or group, and the Signal to Noise ratio and reliability.

- **Reviewer 6**: All attribution models are subject to differences of opinon but eventually, once a decision has been made, it must be reasonable and tested. The "up-to-3" provider model chosen here, though certainly subject to exceptions, allows for flexibility and applicability to allow for discharging provider (though not necessarily admitting provider), pleurality of care provider during the admission (who may nor may not be admitting, discharging, or the outpatient provider), and pleurality of care provider as an outpatient (who may be none of these either). This same model is also applied to clinician groups according the the applicable TIN. While there are many possible exceptions to this attribution model, it was the one tested with moderate to substantial reliability results. Clinicians and clinician groups were explicitly and appropriately excluded from the denominator for a number of well-delineated reasons.
- **Reviewer 5**: Details regarding s.8 – exclusions are required.
- Code listings and/or CCS listings for exclusions should be supplied.
- **Reviewer 3**: The risk model and attribution strategy are a little bit complex, but the MIF is very clear. 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 <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

☐ Yes ☐ No

N/A. Score level testing was conducted.

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

Submission document: Testing attachment, section 2a2.2

- **Reviewer 1:** Signal to noise ratio for inter-clinician comparison.
- **Reviewer 2:**Signal-to-noise; appropriate methodology.
- **Reviewer 6**: The testing utilized a split half methodology, with two years of administrative data. Dataset A1 was the development sample, comprising over 3 million admissions, 550,000 clinicians, and 117,000 clinician groups. Dataset A2 was the validation sample over the same timeframe with over 3 million admissions, 550,000 clinicians, and 117,000 clinician groups. A second validation sample set, B, was used for the second year of testing. A1, A2, and B were used for reliability testing, A1 and A2 were used for exclusion testing, A1, A2, and B were used for risk adjustment testing, and A1 and A2 were used for difference in performance testing.
- **Reviewer 5**: Method is appropriate, but only reported results for clinician/groups with at least 25 observations. I did not see that in the measure specs – may have missed it.
- **Reviewer 3**: Developer use Adams reliability measure for each cohort, requiring a minimum of 25 cases per unit of attribution. Makes sense.
- **Reviewer 4**: Appropriate

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

Submission document: Testing attachment, section 2a2.3

- **Reviewer 1:** Sample size is adequate, or at least what will be observed in practice. My principle concern is that the S-to-N ratios, characterized as moderate for clinicians and substantial for clinical groups are too low to assure correct classification of clinicians or clinician groups as high or low performing.
- **Reviewer 5**: Moderate
- **Reviewer 2:** Average reliability at the group level, reliability was 0.55, 0.65, 0.47, 0.57, 0.45, and 0.54.
- **Reviewer 4**:
- Mean Signal to noise ratio at clinical group level
    - Cardiorespiratory Neurology Medicine Cardiovascular Surgical Average
    - Mean 0.55 0.65 0.47 0.57 0.45 0.54
- The developers conclude reliability is "moderate" to "substantial". I would argue that for clinical groups, the reliability for Medicine and Surgical cohorts of <.50 is at best moderate.
- **Reviewer 3:** At the provider group level (TIN) the scores are lower (0.45 for Surgical up to 0.65 for neurology). It would be helpful to see the distribution of reliability scores, including median, rather than just the mean.
- **Reviewer 6**: Reliabilty testing was performed at both the clinical level and the clinician group level. The former demonstrated mean signal-to-noise ratios of 0.55 to 0.77 across various clinical specialties, while the latter demonstrated mean signal-to-noise ratios of 0.45 to 0.65 across various clinical group specialties. These results are between moderate and substantial agreement.
- Entities, whether the clinician or clinician group, had to have at least 25 attributed index admissions to be included.

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 <u>all</u> testing results):

Methods Panel members varied on their response to this question. Responses ranged from moderte to low but the panel ultimately came to consensus that the measure demonstrated moderate reliability.

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - **Reviewer 1:**I do not consider a S-to-N ratio of <.6 at the clinician level or clinical group level to be sufficiently reliable to prevent misclassification in use for payment, which is the intended use of this measure.
    - I would like to see additional tests of the consistency of the results across samples, test-retest in split samples or across years, proportion of times clinicians or clinician groups remain classified as above, no different from, or with poorer performance than others.
    - **Reviewer 2:**ICCs suggest moderate reliability.
    - **Reviewer 6**: Extensive and appropriate reliability testing was performed on a very large dataset within the assumptions made about attribution. It is interesting that incorporation of the social risk factors did not significantly improve the model performance. Measure score reliability ranged from moderate to substantial by Landis and Koch standards.
    - **Reviewer 3**: Given the low reliability scores, Moderate is the highest this measure can rate.
    - **Reviewer 4**: Interpret signal to noise values as moderate at best for some cohorts.

VALIDITY: SUMMARY

<u>Validity</u>

- Validity testing was performed for the measure score via empirical validity testing and face validity.
- Empirical Validity Testing
    - For empirical validation, the developer assessed how well the measure correlated with hospital quality as determined by the CMS Hospital Overall Star Ratings and Hospital Star Ratings readmission domain scores.
    - External validity results suggest that eligible clinician groups' risk adjusted readmission rates go down with increasing overall hospital quality Star Rating and with increasing quintile of the Star Rating readmission quality score.
- Face Validity
    - Seventeen Technical Expert Panel members were asked to assess the face validity of the final measure specification by confidentially responding to two questions:
        - The risk-standardized readmission rates obtained from the MIPS HWR measure as specified:
            - Are valid and useful measures of MIPS EC and MIPS EC group quality of care?
            - Will provide MIPS ECs and MIPS EC groups with information that can be used to improve their quality of care?
        - TEP members were asked to report their agreement with each statement on a 6-point scale, representing a range from "strongly disagree" to "strongly agree."
    - The majority of respondents, 12/17 or 70 percent, agreed that the HWR measure scores were valid and useful, and the same proportion agreed that the measure would provide information that could be used to improve the quality of care. Among those who disagreed, the primary concern was that factors which led to increased risk of readmission were beyond the control of any single eligible clinician or clinician group.
    - Overall, SMP subgroup members indicated the approach to validity testing was appropriate and that results demonstrated moderate validity.
    - This measure uses a statistical risk model with 32 + variable number of condition category (CC) risk factors.

- The developer notes that the measure estimates clinician-level 30-day all-cause risk-adjusted readmission rates using reliability-adjusted observed to expected ratios. They explain that the approach first models data at the patient level to construct clinician group level observed to expected ratios, and then adjusts these ratios to account for variance in patient outcomes within and between clinician groups.
- As the intention was to harmonize this measure with the existing hospital level measure (NQF #1789) the same risk factors were adopted.
- The developer provided a conceptual rationale for how social risk factors could influence readmissions. The developer performed a literature review and categorized their findings into three domains of social risk factors that have been examined: 1) patient-level variables, 2) neighborhood/community-level variables, and 3) hospital-level variables. Patient-level variables describe characteristics of individual patients and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level. The developer identified at least four pathways through which social risk could influence readmission rates: 1) relationship with health at admission, 2) use of low-quality hospitals, 3) Differential care within a hospital, and 4) influence on readmission risk outside of hospital quality and health status.
- Based on the findings of the conceptual model and available data sources, the developer examined the impact of two social risk factors: dual eligible status and AHRQ SES index. The developer examined the relationship between these variables and the outcome and examined the incremental effect in a multivariable model. The developer also examined the extent to which the addition of this variables improved model performance or changed hospital results
- The developer found that the patient-level observed hospital wide readmission rate is higher for dual-eligible patients, 19.51%, compared with 14.63% for all other patients. Similarly, the readmission rate for patients in the lowest SES quartile by AHRQ Index was 17.42% compared with 14.90% for all other patients. The developer then examined the strength and significance of the SDS variables in each of the five specialties cohort multivariable models and found a modest effect size. The developer found that the addition of any of these variables into the model has little to no effect on eligible clinician or eligible clinician group performance.
- The developer calculated three summary statistics to assess model performance:
    - Area under the receiver operating characteristic (ROC) curve (the c-statistic)
    - Discrimination – Predictive ability
    - Calibration value of close to zero at one end and close to 1 on the other end
- The calculated c-statistics ranged from 0.63 to 0.71 across specialty cohorts and datasets.
- The calibration values which are consistently of close to 0 at one end and close to 1 for all specialty cohorts and datasets
- SMP subgroup members raised some concerns that social risk factors were excluded from the risk model given the effect size and the potential for negative consequences on access to care if this measure is not adequately risk adjusted. Scientific Methods Panel members suggested the developer examine other clinical variables that could underlie disparities such as frailty or functional status.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

- **Reviewer 1:**Validity of the readmission measure as constructed was considered in the endorsement of the hospital-level version of the measure. Exclusions appear appropriate, and numbers excluded reasonable.

- **Reviewer 2:**These were shown at the hospital level, which is inconsistent with how the measure would be used. The impact of exclusions needs to be shown at the clinician and group level to be consistent with measure. specifications. I don't disagree with the exclusions themselves, however.
- **Reviewer 6**: Meaure exclusions are specifically delineated and are appropriate. Testing of each exclusion criteria was performed and accounted a range of 0.3% of all index admissions (for cancer) to 2.45% for unmatched patients. Analysis and conclusions from the results is provided.
- **Reviewer 5**: Developer did not include enough detail to assess – unclear how 'admitted for medical treatment of cancer' or psych dx are identified
- **Reviewer 3**: I'm not clear on why cases with discharge AMA are excluded . Taking an intent to treat view, these represent care that could have potentially been better managed. Also, transfers are a concern – on the one hand, excluding these cases makes sense from a 'span of control' point of view. Those in the receiving hospital did have to control over early care. That said, cases should be free to move through the system to the best level and setting of care. Adequate transfer of information between hospitals should be the rule, not the exception. These two issues aside, less than 10 percent of cases are dropped – that is excellent.
- **Reviewer 4**: None.

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

Submission document: Testing attachment, section 2b4.

- **Reviewer 1:**The reported range of variation is 10^{th}-90^{th} percentiles is 12.6%-18.0% for clinicians, and 13.8-17.1% for clinician groups. If these were persistent differences, they would be meaningful. It is not clear that they are persistent, based on reliability reported.
- Attribution will be an issue in this measure, given it is being concurrently applied to up to 3 physicians – discharge physician of record, predominant inpatient physician, predominant outpatient physician. Not clear how predominant outpatient physician relates to post-acute treatment for hospitalized patients if care is highly specialized and focus of inpatient treating physician. Also not clear how this attribution relates to hospitalists, who will have minimal engagement with patients once the patient is discharged.
- Weighting and comparison group when a patient is attributed to three patients is described and seems plausible, but would like more discussion of how often the triad is the same over patients.
- **Reviewer 2:**No concerns
- **Reviewer 6**: Performance testing was done and identified the 19,000 significant utliers in the 170,000 clinicians and also the 6,400 outliers in the 55,000 clinician groups. The overall RARR was 15.1% for clinicians and 15.3% for clinician groups. More than 10% in both categories were outside of the expected performance range.
- **Reviewer 3**: No concerns.
- **Reviewer 4**: None.

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.

- **Reviewer 2:**No concerns
- **Reviewer 6**: Not applicable
- **Reviewer 3**: NA.
- **Reviewer 1:** NA

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

Submission document: Testing attachment, section 2b6.

- **Reviewer 2:**No concerns
- **Reviewer 6**: See above. ..Unmatched data accounted for 2.45% and a lack of attribution assignment for 2.13% of the data
- **Reviewer 3**: NA.
- **Reviewer 1:** NA

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No **Reviewer 1:**NA

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?

Methods Panel members varied on their response to this question. Some responded it was appropriately risk-adjusted while others disagreed.

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

- **Reviewer 1:**The risk adjustment and its analysis is well done. Adequate C stats, good correlation of predicted and actual rates of readmission by quintile. Good discussion of SES risk adjustment, and consideration of two risk adjusters: dual eligibility and census block SES.
- **Reviewer 2:**Basic model with 30 clinical variables. No prior utilization or functional variables.
- Dual effect was significant and sizeable, with odds ratios ranging from 1.07 to 1.16 after adjusting for clinical variables. Dual status was not included because the developers did not feel that adding it made a significant difference in terms of model performance and the relative performance of

many clinicians and groups was unchaged. It seems that for the groups whose readmission rates were changing by more than 2%, it might still be a significant change (and these would likely be the groups we would least likely want to give an incentive to avoid caring for poor indivduals).

- If CMS is committed to not adjusting for social risk factors, it is incumbent upon measure developers to demonstrate that they have done adequate risk adjustment to account for the medical factors that they think might underlie these very striking disparities, such that we can be certain that any signal that remains is a quality signal and not the result of unmeasured confounding. To that end, exploring markers of frailty, functional dependence, behavioral / mental health diagnoses, etc. that might influence revisits – even if they are not ones that a technical expert panel might suggest – would be of utility. Another thing that could be done to reassure us that the dual effect is a quality signal rather than residual confounding would be to explore the reasons duals return – related versus unrelated, timing, etc.
- **Reviewer 6**: This is a claims based measure based on a model for publicly reported outcomes articulated in the AHA Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes". Four other condition specific claims-based readmission models use CMS-CC's model to group codes ito risk adjustment variables and are validated. Variable selection was performed from 30 vaiablers drawn from previous readmission models and 11 additional CMS-CC's were deemed relevant to an all-condition measure. Risk variables were excluded that were statistically significant only for a few conditions, those that behaved in clinically incoherent ways, those that were predominantly protective and not clinically relevant, and some previous ones were grouped by category. The service mix adjustment is described with the use of a condition-specific variable for all condition categories with sufficient volume. SES factors were included and examined.
- **Reviewer 5**: Developer justifies the removal of SES variables due to lack of impact on model fit**.** At the same time, the results displayed in 2b3.4b show that the odds ratios are higher for both dual eligible and Low SES variables. Developer should go one step further and show that other variables are indeed correlated with SES and acting as a proxy.
- **Reviewer 3**: C-statistics and calibration statistics look good. I particularly like the use to two different time periods (2015-2016 and 2016-2017). The Calibration plots are very helpful. Overall, seems like a good risk adjustment strategy.
- **Reviewer 4**: Odds ratio results for both SES variables tested showed statistically significant and in some case quite large impact on readmission rates (e.g., 10% to 16% higher readmission rates in many cases). This was fully backed up by the literature reviewed. However, developers decided NOT to adjust for SES because it had little impact on c-statistic or model performance. Again, do not think that is good rationale to exclude. It may be very impactful to some clinicians and groups serving large populations of disadvantaged patients, thereby penalizing and limiting access to care for high need populations.

	Dual Eligible	Low SES
• Cohort	Odds Ratio	Odds Ratio
• CARDIORESPIRATORY	1.12	1.10
• CV	1.13	1.09
• MEDICINE	1.07	1.05
• NEUROLOGY	1.10	1.07
• SURGICAL	1.16	1.08

VALIDITY: TESTING

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

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

☒ Face validity

☒ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

- **Reviewer 1:**TEP, correlation with hospital star ratings based on readmission measure. Both seem reasonable.
- **Reviewer 2:**c-statistic and calibration
- **Reviewer 6**: Face validity. Three TEP conference calls were held to discuss whether the provision of risk-standardized readmission rates were valid and useful measures of EC and EC group quality of care, and would that information be helpful for quality improvement. Surveying was performed from "strongly agree" to "strongly disagree". 70% of the TEP agreed that the measure scores were valid and useful. It is interesting that they also identified that a multiple provider attribution model were be superior to a single attribution model.
- **Reviewer 5**: Comparison to star rankings is weak – no statistical testing of the relationship between to two measures (or lack of relationship). ANOVA would have been easy to complete here.
- **Reviewer 3**: Clever approach to empirical validity testing with hospital compare as an external data source. The box and whisker plot on page 14 of the MIF is particularly helpful for visualizing the trend in scores.
- **Reviewer 4:** Appropriate.

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

Submission document: Testing attachment, section 2b2.3

- **Reviewer 1:**TEP was supportive but not overwhelmingly of the measure. 70% viewed it as valid and useful. No indication of how strong or weak the positive views were.
    - Concordance of clinician rankings and hospital rankings existed, but large overlap across all 5 hospital rankings.
- **Reviewer 2:** c-statistic was 0.63-0.70 across cohorts. Predictive ability was reasonable across cohorts
- **Reviewer 6**: C-statistic for discrimination was 0.636 to 0.71 for datasets A1,A2, and B. Calibration statistics are provided and were close to one for all specialty cohorts and datasets. Risk Decile Plots indicated excellent discrimination of the model and good predictive ability.
- **Reviewer 5**: Comparison to star rankings is weak – no statistical testing of the relationship between to two measures (or lack of relationship). ANOVA would have been easy to complete here.
- **Reviewer 3**: Interesting results – I would have expected a more pronounced difference at the upper end of the hospital compare quality distribution. This just goes to show how hard it is to 'move' readmissions. The polling of the TEP is interesting, but it does not really forecast if and how providers will use this measure in operations – this is something it would be valuable to follow-up on over time.

- **Reviewer 4:** Empirical validity testing assessed how well measure rates correlated with hospital quality and found good correlation. Face validity tested using TEP.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

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

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

☐ **Yes**

☐ **No**

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

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

Methods panel members varied on their response to this question. Responses ranged from high to low.

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

- **Reviewer 1:**I would like more information on the strength of the TEP ratings among those judging the measure valid and useful.
    - It would be useful to have more discussion of the attribution approach regarding hospitalists and general physicians.
- **Reviewer 2:** Concerned with handling of dual status.
- **Reviewer 3**: The risk standardized score is modestly associated with the hospital compare measures – although hospital compare is not a gold standard per se, I thought there would be a stronger relationship. Since claims are designed for payment, it does seem hard to get a high validity outcome measure from this source.
- **Reviewer 4**: Overall results show consistently high validity of measure rates.
- **Reviewer 6**: The model does show performance differences and does have predictive ability. I was a little surprised at the lack of a demonstated effect of the inclusion of SES variables in the model. The given explanation is that the clinical factors included in the model dominated the SES variables.

ADDITIONAL RECOMMENDATIONS

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

- **Reviewer 1:**See notes above regarding:

    - Attribution
    - TEP member scoring
    - Concern about S-to-N scores and desire for more detailed analysis about stability of rankings in different samples of data.
- **Reviewer 2:** Yes – lack of adjustment for dual status despite strong relationships. In this case it seems that adjusting for dual status would be appropriate.

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

Reliability – Specifications

- Being used and clearly defined
- I would like greater clarity that transportation, access to healthcare, including availability of providers and community health programs, food insecurity, education and economic challenges related to prescription costs have been considered in risk-adjustment.
- Measure specs are adequate
- Which data elements, if any, are not clearly defined? none Which codes with descriptors, if any, are not provided? none 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? none What concerns do you have about the likelihood that this measure can be consistently implemented? none other than wrong claims
- Methodology for reliability testing appears appropriate. Of note, social risk factors were excluded and could potentially impact consistency
- Data are adequate; no concerns
- The data elements, steps in the logic algorithm and other specifications are clearly defined and set forth. I was unable to link to the codes. I have some concern about the three prong attribution model described. It is complex and the Primary Inpatient Care provider may be an assigned hospitalist who has no connection to the patient's pre-hospitalization or post discharge care. Without some way to account for when a hospitalist is the PIC or DC, the objectives of this measure may be blurred and a potential source of inconsistency between clinician groups develop.
- No concerns
- The description is clear. Why are readmissions, and not excess days in acute care (including ED and observation) part of the numerator? Would seem to warrant consideration given the controversy over the impact of HRRP and Sabbatini et al PMID: 29847758
- Primary inpatient care provider definition is skewed to billable charges. When IP diagnostics are completed and patients are "ruled out" the diagnostic physician of record might be the Highest fees? this would be especially true when the "admit" definition includes ED, short stay and observation stays. Or is the definition of the admit changed to accomodate this? If the diagnostic physician is considered accountable for readmit, the workflow and care does not allow for interventions as the patient is "ruled out". I would indicate low reliability as a result.
- Provider and primary care attributions would be difficult to implement consistently.
- It would be helpful to have more information on risk/case-mix adjustment.

Reliability – Testing

- No concerns
- I have no concerns.

- Lingering concerns re: social risk and as the SMP noted, geriatric medicine issues such as adverse selection for a frailer population.
- no
- Concern that the social risk factors were excluded.
- No
- The signal to noise reliability testing scores showed only moderate agreement for the medicine and surgical cohorts except at the higher case volume numbers. I am concerned about reliability at lower case volumes clinical groups.
- None
- again, more information about the case and outcome volume at the clinician/ clinician group level would help
- as noted above; I would further add that measurement should drive quality improvement. Those involved in improvement (providers) should know organically that they are responsible for care and readmissions. Claims data as a source, and providers defined attributed by fees are not naturally known in the care process. Attribution to dollar amounts means nobody in care knows who is accountable for this measure while the patient is in care.
- I do have concerns about this measure due to its dependence on provider attribution. In addition, identifiyng who is a primary care provider is even harder for the elderly. Is a heart specialist a primary care provider if that specialist giving the most care to the patient?
- None currently.

- **Validity – Testing**
  - No concerns
  - Moderate validity was demonstrated and the results seemed reasonable to this reviewer.
  - No
  - no
  - Appears adequate
  - No
  - I have concerns about conclusions drawn from empirical validity testing using the Star Rating System since it has come under criticism from hospital associations and is in the process of being updated by CMS. Also, it is not clear to me that those results would apply to a clinician group level measure since they are determined for the hospital level.
  - NONE
  - no
  - I share this concern from TEP: Attribution will be an issue in this measure, given it is being concurrently applied to up to 3 physicians – discharge physician of record, predominant inpatient physician, predominant outpatient physician. Not clear how predominant outpatient physician relates to post-acute treatment for hospitalized patients if care is highly specialized and focus of inpatient treating physician. Also not clear how this attribution relates to hospitalists, who will have minimal engagement with patients once the patient is discharged.
  - Yes I do have concerns due to the exclusion of claims with missing provider information. Also this measure doe not address the probablty of the competing outcome of death, which has a higher probability in the age 65+ population.
  - The only concern is potentially with changes with risk/case-mix adjustment methodology.

- **Validity – Threats**
  - No concerns
  - This reviewer did not identify any obvious missing data sources that would pose a risk to measure validity.
  - As per the ACO measures, this one may also have issues with attribution.

- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) none 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? unsure 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? yes 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure? no, data seems to be fairly complete
- There is a concern for attribution in the measure.
- No concerns
- 2b4:The RARR median percentage is compared with the 10th and 90th percentile to determine high and poor quality clinician group level results. Another approach to assess meaningful differences about quality might be to assess and compare a small number of the IQR hospital measures that are clinician dependent. 2b5: Not applicable. 2b6: This is a claims-based measure and therefore there are no missing data.
- I have no concerns
- no
- no
- Yes, missing data constitudes a red flag.
- No, not in my opinion.

- **Validity – Other**
- Moderate
- I have commented earlier about concerns that Social Determinants of Health many not have been adequately considered in the risk adjustment. I do not, however, have empirical proof that they weren't.
- Ongoing issues re: social risk adjustment, perhaps dual eligibility should remain in.
- 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? yes How well do social risk factor variables that were available and analyzed align with the conceptual description provided? this needs further clarity Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? difficult to gather this data Was the risk adjustment (case-mix adjustment) appropriately developed and tested? yes Do analyses indicate acceptable results? yes Is an appropriate risk-adjustment strategy included in the measure? yes, to best available data
- Concern about the social risk adjustments
- No concerns
- 2b2: The exclusions are consistent with the available evidence, data availability and/or the stakeholder consensus. There are no groups inappriately excluded from the measure. 2b3: Yes, there is a conceptual relationship between the two potential social risk factor variables considered, dual eligibles and AHRQ SES Index score, and the measure focus. The social risk factors considered align with two of four conceptual framework descriptors delineated, relationship to health at admission and influence of SES on readmission risk outside of hospital quality and health status. The risk adjustment variables considered were available at the start of care. The risk adjustment was appropriately developed and tested. Results indicated that the SES risk factors did not improve model performance and were not included in the final risk model. CMS CC's were used to case-mix adjust and were felt to capture much of the risk related to low SES. The risk adjustment strategy is very complex and seems to be appropriate with the question of the wisdom of not including SES factors shown to have some association with readmission risk.
- No concerns

- I am still concerned about the lack of social risk adjustment in a measure that will be used for payment, though I understand the rationale for excluding. It's possible that safety net hospitals will also code comorbid conditions less aggressively, which could negatively impact their risk adjustment.
- as considered for dual eligible
- Yes, I think so.
- It would be helpful to have more information on social risk factor variables and other potential sources of data collection.

Criterion 3. Feasbility

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.

- This measure uses claims data that has been shown to be operationalizable, however, the measure is not yet in use.
- There are no fees, licensing, or requirements to use the measure.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Is the data collection strategy ready to be put into operational use?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

- No concerns
- The data collection strategy seems perfectly reasonable. I have no concerns about feasibility.
- Claims data, shouldn't be an issue
- Which of the required data elements are not routinely generated and used during care delivery? none per my expertise and knowledge Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? depends on EHR some hrs do not have clinician claims What are your concerns about how the data collection strategy can be put into operational use? claims (revenue cycle module) can be not part of electronic record and be manual through billing companies
- Uses claims data
- No concerns
- This is a claims-based measure and all data elements are routinely available, but not routinely generated and used during care delivery.
- The data elements are from claims data, and should be fairly standard and reliable.
- ok
- claims data. However as noted above, if charges are the only basis for attribution, nobody will know in the moment of care, who is responsible? And even retrospectively how would a improvement plan be initiated when highest charges are the basis for care accountability?
- The dependency of this measure on the provider data is a concern.
- Data on functional status would be valuable but is not uniformly collected, however some information may be availabl in electronic form.

Criterion 4: Usability and Use

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

Merit-based Incentive Payment System (MIPS). MIPS consolidated Medicare's existing incentive and quality reporting programs for clinicians into a single program. MIPS makes positive and negative payment adjustments for Eligible Clinicians (ECs) based on performance in four categories:

- Quality
- Cost
- Advancing care information
- Improvement activities

To meet the quality component of the program, individual ECs or groups of ECs choose six measures to report to CMS. One of these measures must be an outcome measure or other high-priority measure. Clinicians can also choose to report a specialty measure set.

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

Feedback on the measure by those being measured or others

- The developer did not provide any information on the measure vetting.

Additional Feedback:

- The developer did not provide any information on feedback provided on the measure.

Questions for the Committee:

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

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

RATIONALE:

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

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

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

Improvement results

- The developer notes that this is a new measure and there is no information available on performance improvement. This measure is not currently used in a program, but a primary goal of the measure is to provide information necessary to implement focused quality improvement efforts. Once the measure is implemented, the developer plans to examine trends in improvements by comparing RSRR over time.

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 did not note any unexpected findings from measure testing.

Potential harms

- The developer did not provide any information on potential harms.

Additional Feedback: N/A

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**

RATIONALE:

- No information was provided.

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

Use

- yes
- The provider appears to have provided adequate opportunity for feedback. Previous concerns about the facility-based focus were considered and reflected in this resubmission.
- Not currently being used,
- 4a1. Use - Accountability and Transparency: How is the measure being publicly reported? star?? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? not currently For maintenance measures - which accountability applications is the measure being used for? n./a For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? needs

further discussion. 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? N/A Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? N/A-NO Has this feedback has been considered when changes are incorporated into the measure? N/A

- Measure is not in use yet.
- Uncertain
- 4a1: Not applicable. 4a2: This is a new measure and not currently in use.
- I have no concerns. It appears the developer didn't report vetting information, but measurement of readmission is a fairly common thing.
- where the RARR numbers vary, there doesn't seem to be a large magnitude difference between 1 and 5 star providers based on page 52 figures, so I'm not sure that is the ideal way to report
- Yes
- Not clear
- The measure is being publicly reported and serves to promot accountability. For conditions where Medicare fee-for-service is not the primary payer, publicly reported data may be misleading without clear guidance to consumers.

Usability

- No harm identified by the developer
- The provider didn't offer any information regarding potential harm from implementing this measure. This reviewer wasn't able to identify any either.
- Not currently being used. Considering that this is a measure focused on clinical groups who care for a significant proportion of MCR beneficiaries (Otherwise they would not be MIPS eligible), if as with other MCR measures the outcomes are used to penalize lower performing groups, there may be an enhanced push towards refusing to take MCR, at least in the outpatient groups. Do not think this will pertain to the inpatient groups, as they will likely have no choice. On the other hand, groups might choose to use financial incentives to improve clinical infrastructure to improve performance.
- 2. 4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? in mips program - the payment based carrots and sticks, we hope to influence physician involvement in readmission reduction 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? NO, needs more explanation 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them. measure does have potential to use observation status or other level of care more. physicians might not want to take care of couples patients with high readmission risk
- New measure and no information available on performance improvement with the measure
- None apparent
- 4b1: Feedback of data to clinician groups in a timely fashion and implementation of practice transformation techniques to make care more patient-centered and coordinated. examining trends in improvement by following RSRR over time. There is no discussion of use of the measure for value based purchasing in the MIPS program. 4b2: This is a new measure. No known unintended consequences. There is no discussion if there have been any noted unintended consequences for the hospital measure.
- This measure would help in performance improvement and I'm not aware of any harms of this measurement

- many patients don't have the ability to select their Primary Inpatient Provider or Discharge Clinician. Payers may use the reporting to steer patients to clinician groups in ways that could adversely impact patient out-of-pocket costs
- As noted above, attribution of care by charges are not consistent with accountable provider
- This measure could potentially reward facilities that do not report complete provider information. Or it could also reward providers and facilities with high mortality rates, since those with high mortality rates would likely to have lower rehospitalization rates. It might be best if this measure is coupled with a 30 day mortality measure.
- There is a credible rationale that the results could be used to further the goal of high-quality, efficient health care. Without a disclaimer, populations may think that Medicare data applies to them when it does not.

Criterion 5: Related and Competing Measures

Related or competing measures

- NQF #1789, Hospital Wide All-Cause Unplanned Readmission Measure.

Harmonization

- The developer notes that this measure is aligned with #1789 but the attribution is to a clinician or clinician group rather than a facility.

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

- Yes, harmonized
- There are no competing or related measures in this focus area about which I am aware.
- Similar to the hospital wide readmission measure.
- no
- Measure is aligned with NQF #1789 however the attribution is to a clinician or clinician group versus a facility
- None apparent
- This measure is a re-specification of NQF #1789, Hospital Wide All-Cause Unplanned Readmission Measure. Measure specifications are harmonized to the extent possible. The attribution is not to the hospital facility but to potentially three clinician groups to create incentives for shared accountability for care coordination and patient readmissions.
- There are other readmission measures, but at the hospital/health system level; this measure would be complementary to that
- no
- If admit definition should be harmonized or specifically called out, so that diagnostic providers are not accountable providers, OR admission is defined as in the original measures. (not include obs stays, short stays, ED)
- This measure is a varient of the original hospital based version.
- Readmissions and successful discharge to home measures should be harmonized.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/29/2020

- **Of the one NQF member who have submitted a support/non-support choice:**
  - One comment and one expression of non-support have been submitted.

Comment: The American Medication Association (AMA) appreciates the updated information provided by the developer on this measure but we continue to believe that the evidence and testing provided do not meet the NQF Measure Evaluation Criteria.

The additional information within the evidence submission outlining the justification for attribution to the three types of clinician groups relies on general statements and only two additional studies are cited specific to attribution to the discharging clinician. One article focuses on individuals with a diagnosis of heart failure and while it is a meta-analysis of multiple studies, it does not directly demonstrate that clinician action is what leads to decreased readmission rates. The second study is one that shows that the use of a decision support tool by physicians can assist in better discharge processes and ultimately reduced readmission rates. While this finding is encouraging, it is not broadly applicable since the intervention was only implemented across four medical units in one urban, university medical center. Interestingly, while the researchers were able to reduce referral or high-risk patients readmissions, the rates (even when improved) are around 17%, which is similar to the current performance data provided in 1b. Performance Gap. Therefore, raising a question that we have asked and highlighted in previous reviews of the hospital level measure (NQF 1789) on whether there are any additional reductions in rates to be gained.

In addition, the measure score reliability across the 5 specialty cohorts continues to be below a minimum acceptable threshold of 0.7 when a case minimum of 25 patients is applied. The results continue to remain less than optimal when a minimum sample of 200 patients is applied.

The AMA believes that this additional information, while helpful, does not alleviate any of our concerns and encourage the Standing Committee to not recommend the measure for endorsement.

-

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

MIPSHWREvidenceForm10.31.19_v1.0.docx,MIPSHWRIntentToSubmit10.31.19_v1.0.docx

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3495

Measure Title: Hospital-Wide All-Cause Unplanned Readmission Measure for MIPS Eligible Clinicians

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: Click here to enter a date

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: 30-day all-cause readmission

☐ Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

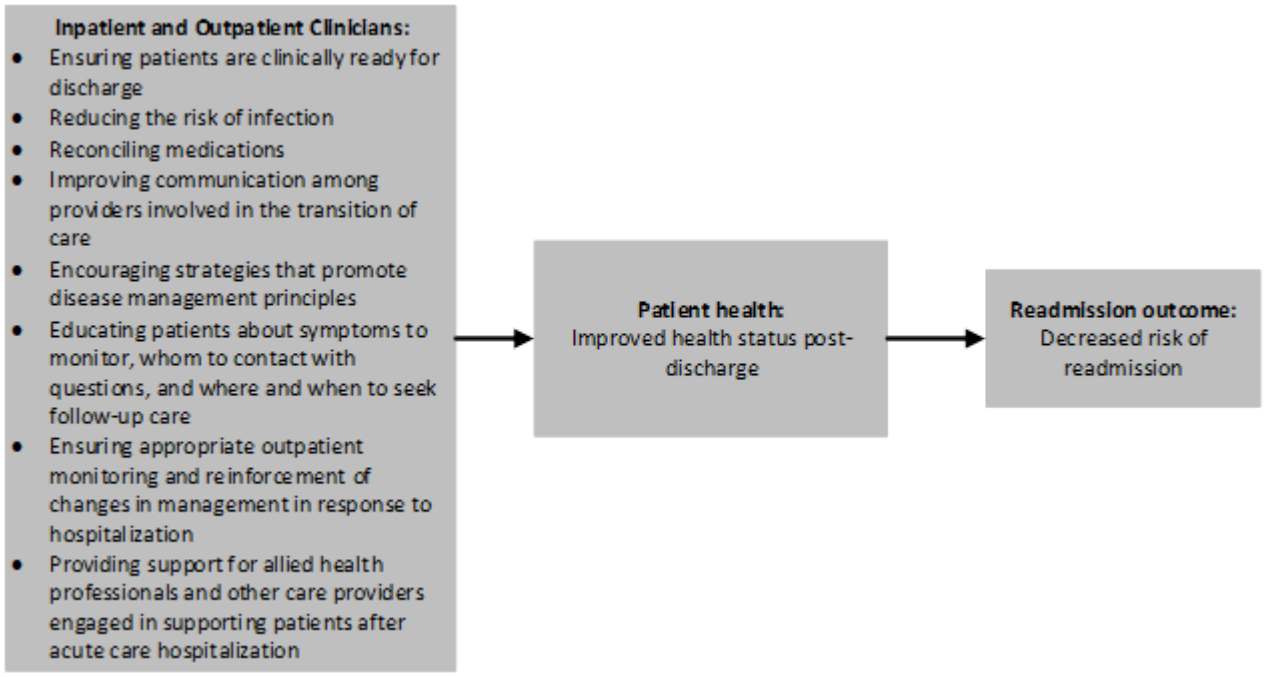
☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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.



The Hospital-Wide All-Cause Unplanned Readmission Measure is an adaptation of an existing, publicly reported measure for hospitals. The goal of this measure is to improve patient outcomes by providing patients and clinicians with information about clinician-group level, risk-standardized readmission rates of unplanned, all-cause readmission after admission for any eligible condition within 30 days of hospital discharge. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This readmission measure was developed to identify clinician groups, whose performance is better or worse than would be expected based on their patient case-mix, and therefore promote quality improvement and better inform consumers about care quality.

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.

The diagram above indicates some of the many care processes that can influence readmission risk. In general, randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates: quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; pre-discharge assessment; and coordination of care after discharge. Evidence that hospitals have been able to reduce readmission rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect readmission rates. Successful

randomized trials have reduced 30-day readmission rates by 20-40% [1-11]. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions, and applying lessons learned from clinical trials. Several have been notably successful in reducing readmissions. The strongest evidence supporting the efficacy of improved discharge processes and enhanced care at transitions is a randomized controlled trial by the Project RED (Re-Engineered Discharge) intervention, in which a nurse was assigned to each patient as a discharge advocate, responsible for patient education, follow-up, medication reconciliation, and preparing individualized discharge instructions sent to the patient's primary care provider additionally there was a follow-up phone call from a pharmacist within four days of discharge, which demonstrated a 30% reduction in 30-day readmissions [1].

Studies have shown readmissions to be related to quality of care, and that interventions have been able to reduce 30-day readmission rates, it is reasonable to consider an all-condition readmission rate as a quality measure. [16-26] The variation in readmission rates across hospitals indicates room for quality improvement; targeted efforts to reduce these readmissions could result in better patient care and potential cost savings.

As noted, many of the strategies and best practices used by hospitals to reduce risk of readmissions can also be adopted by groups of clinicians, which clinician groups can influence during the inpatient stay. [27-29] Medication reconciliation, discharge instructions, and outpatient follow up are examples of proven interventions which clinicians can influence during the inpatient stay and during care transitions [12]. To support this model, we found that risk adjusted readmission rates vary from 7% to 25% for eligible clinician groups for 2015 through 2016. Both the high baseline rate and the variability across clinician groups speak to the need for a quality measure to prompt greater care improvement.

Specifically, the measure considers joint attribution for up to three clinician groups or practices that provide care for patients inside and outside of the hospital prior to discharge and are therefore in position to influence patients' risk of readmission. This approach to joint attribution was supported by an extensive literature review and stakeholder input from patients, providers, and payers.

The measure attributes each readmission to:

1. the clinician group of the Primary Inpatient Clinician (PIC),
2. the clinician group of the Discharge Clinician (DC), and
3. the clinician group of the Primary Outpatient Clinician (POC).

Though the same group may be identified as filling two or even three of these attribution roles, the readmission is only attributed once. By holding these three clinical groups jointly accountable, the measure aims to incentivize collaboration of care across inpatient and outpatient settings. Below we list the general responsibilities of each of these clinician groups and how their choices and patterns of care have been shown to influence readmissions. We heard clearly from stakeholders that readmission is a complex clinical outcome requiring multidisciplinary and collaborative care to reduce risk and optimize health outcomes. Clinicians are critical components of quality improvement, whether as change agents and influencers of health systems or key stakeholders whose acceptance is required for any successful or long-lasting improvements, which is why CMS chose to implement readmission measurement into the MIPS program.

The PIC group is defined as the clinician group for the PIC for a given patient - that is, the clinician who billed the most charges for the patient during their hospital say. Such clinicians are most likely responsible for ensuring relevant medical problems are addressed in the inpatient setting, reducing the chance patients will return to the hospital with unresolved medical issues. For example, a medical specialty PIC is the clinician likely seeing a patient on a regular basis while a surgical specialty PIC likely performed a significant surgery for the patient. These clinicians, and the clinician groups they are part of, represent one of the three clinician groups to which this measure attributes readmissions. They are likely to make decisions about what medications the patient needs and what other specialties or providers should be involved in the patient's hospital care. Studies

have shown that selection of affordable medications with favorable side effect profiles and dosing schedules that are easiest to adhere to, such as daily or twice daily dosing, predict adherence and lower readmission.

The Discharge Clinician (DC) group is defined as the clinician group for the DC for a given patient - that is, the clinician transitioning the patient from inpatient to outpatient care. The Discharge Clinician (DC) is most responsible for preparing a patient for discharge. Responsibilities for the DC include determining that the patient is, in fact, clinically appropriate to leave the hospital and ensuring they understand their medical condition and treatments. The DC can also directly support and influence other providers and staff to support critical actions related to discharge: providing hard copies of discharge instructions in a language the patient can understand, ensuring the instructions are concordant with what was communicated to the patient by the PIC and other clinicians during the inpatient stay, and verbal explanation of the discharge instructions including management of medications, referrals to outpatient specialists or therapy, and lifestyle modifications, all of which ensure the patient adheres to the plan of care upon discharge from the hospital [30, 31]. Studies have shown that interventions focused on inpatient clinicians such as the PIC and DC can strengthen discharge systems by making phone calls, scheduling appointments, and providing information on the transition to a Primary Care Provider – these have all been shown to improve patient access to a clinician once they are discharged from the hospital and reduce readmission.

The Primary Outpatient Clinician (POC) group is defined as the clinician group for the POC for a given patient - that is, the clinician with the greatest number of claims for primary care during the 12 months prior to the hospital admission ate. This clinician can reduce the chance that a patient will be readmitted by having open access and ensuring available appointments for the patient within 30 days of discharge.

Further, this measure complements the hospital measure as a proportion of clinicians have very different performance quality than the institutions in which they see patients; we found that adjusting for hospital overall quality (as measured by CMS's hospital Star Rating score) shifted 68 clinician groups from either first to fifth quintile or fifth to first quintile of performance. The measure presented in this application provides a transparent reflection of these discordances to further support quality improvement.

This evidence - that readmissions within 30 days have been shown to be related to quality of care; that interventions, especially those utilizing clinicians, have been able to reduce 30-day readmission rates for a variety of specific conditions; and the observed high and variable clinician-level readmission rates – all support the logic model in Section 1a.2 and indicate the potential value of clinician group level readmission for driving improvements in care.

References:

1. Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, et al. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med 2009;150(3):178-87.
2. Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention. J Am Geriatr Soc 2004;52(11):1817-25.
3. Courtney M, Edwards H, Chang A, Parker A, Finlayson K, Hamilton K. Fewer emergency readmissions and better quality of life for older adults at risk of hospital readmission: a randomized controlled trial to determine the effectiveness of a 24-week exercise and telephone follow-up program. J Am Geriatr Soc 2009;57(3):395-402.
4. Garasen H, Windspoll R, Johnsen R. Intermediate care at a community hospital as an alternative to prolonged general hospital care for elderly patients: a randomised controlled trial. BMC Public Health 2007;7:68.

5. Koehler BE, Richter KM, Youngblood L, Cohen BA, Prengler ID, Cheng D, et al. Reduction of 30-day postdischarge hospital readmission or emergency department (ED) visit rates in high-risk elderly medical patients through delivery of a targeted care bundle. J Hosp Med 2009;4(4):211-218.
6. Mistiaen P, Francke AL, Poot E. Interventions aimed at reducing problems in adult patients discharged from hospital to home: a systematic metareview. BMC Health Serv Res 2007;7:47.
7. Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M. Comprehensive discharge planning for the hospitalized elderly. A randomized clinical trial. Ann Intern Med 1994;120(12):999-1006.
8. Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. Jama 1999;281(7):613-20.
9. van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med 2002;17(3):186-92.
10. Weiss M, Yakusheva O, Bobay K. Nurse and patient perceptions of discharge readiness in relation to postdischarge utilization. Med Care 2010;48(5):482-6.
11. Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. J Am Coll Cardiol. Jan 2 2002;39(1):83-89.
12. Nelson EA, Maruish ME, Axler JL. Effects of Discharge Planning and Compliance With Outpatient Appointments on Readmission Rates. Psychiatr Serv. July 1 2000;51(7):885-889.
13. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven. New England Journal of Medicine. 1994;331(15):989-995.
14. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine 2000;160(8):1074-81.
15. Medicare Payment Advisory Commission (U.S.). Report to the Congress promoting greater efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.
16. Gil M, Mikaitis DK, Shier G, Johnson TJ, Sims S. Impact of a combined pharmacist and social worker program to reduce hospital readmissions. J Manag Care Pharm. 2013;19(7):558-563.,.
17. Jackson C, DuBard A, Swartz M, et al. Readmission Patterns and Effectiveness of Transitional Care Among Medicaid Patients With Schizophrenia and Medical Comorbidity. N C Med J. 2015;76(4):219-226.
18. Jackson CT, Trygstad TK, DeWalt DA, DuBard CA. Transitional care cut hospital readmissions for North Carolina Medicaid patients with complex chronic conditions. Health Aff (Millwood). 2013;32(8):1407-1415.
19. Kansagara D, Ramsay RS, Labby D, Saha S. Post-discharge intervention in vulnerable, chronically ill patients. Journal of Hospital Medicine. 2012;7(2):124-130.
20. Desai NR, Ross JS, Kwon JY, et al. Association Between Hospital Penalty Status Under the Hospital Readmission Reduction Program and Readmission Rates for Target and Non-Target Conditions. JAMA. 2016;316(24):2647-2656.
21. Wasfy JH, Zigler C, Choirat C, Wang Y, Dominici F, Yeh RW. Readmission rates after passage of the hospital readmissions reduction program: A pre–post analysis. Annals of Internal Medicine. 2017;166(5):324-331.
22. The Medicare Payment Advisory Commision (MEDPAC). Chapter 4: Refining the Hospital Readmissions Reduction Program. Report to the Congress: Medicare and the Health Care Delivery System 2013; http://www.medpac.gov/docs/default-source/reports/jun13_ch04.pdf. Accessed December 18, 2017.
23. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, Observation, and the Hospital Readmissions Reduction Program. New England Journal of Medicine. 2016;374(16):1543-1551.
24. Boccuti C, Casillas G. Aiming for Fewer Hospital U-turns: The Medicare Hospital Readmission Reduction Program. 2017; https://www.kff.org/medicare/issue-brief/aiming-for-fewer-hospital-u-turns-the-medicare-hospital-readmission-reduction-program/. Accessed December 18, 2017.

25. McIlvennan CK, Eapen ZJ, Allen LA. Hospital Readmissions Reduction Program. Circulation. 2015;131(20):1796-1803.
26. Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N. Data Shows Reduction in Medicare Hospital Readmission Rates During 2012. Medicare & Medicaid Research Review. 2013;3(2).
27. Adams, C. J., et al. (2014). "Implementation of the Re-Engineered Discharge (RED) toolkit to decrease all-cause readmission rates at a rural community hospital." Qual Manag Health Care 23(3): 169-177.
28. Sanchez, G. M., et al. (2015). "Revisiting Project Re-Engineered Discharge (RED): The Impact of a Pharmacist Telephone Intervention on Hospital Readmission Rates." Pharmacotherapy 35(9): 805-812.
29. Stauffer, B. D., et al. (2011). "Effectiveness and cost of a transitional care program for heart failure: a prospective study with concurrent controls." Arch Intern Med 171(14): 1238-1243.
30. Bowles KH, Hanlon A, Holland D, Potashnik SL, Topaz M. Impact of discharge planning decision support on time to readmission among older adult medical patients. *Prof Case Manag*. 2014;19(1):29–38. doi:10.1097/01.PCAMA.0000438971.79801.7a
31. Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis [published correction appears in JAMA. 2004 Sep 1;292(9):1022]. *JAMA*. 2004;291(11):1358–1367. doi:10.1001/jama.291.11.1358

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**N/A**

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?	

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 readmission, for any reason, is disruptive to patients and caregivers, costly to the healthcare system, and puts patients at additional risk of hospital-acquired infections and complications. Readmissions are also a major source of patient and family stress and may contribute substantially to loss of functional ability, particularly in older patients. Some readmissions are unavoidable and result from inevitable progression of disease or worsening of chronic conditions. However, readmissions may also result from poor quality of care or inadequate transitional or post-discharge care.

Transitional care includes effective discharge planning, transfer of information at the time of discharge, patient assessment and education, and coordination of care and monitoring in the post-discharge period. Numerous studies have found an association between quality of inpatient or transitional care and early (typically 30-day) readmission rates for a wide range of conditions.1-8 Randomized controlled trials have shown that improvement in the following areas can directly reduce readmission rates: quality of care during the initial admission; improvement in communication with patients, their caregivers, and their clinicians; patient education; pre-discharge assessment; and coordination of care after discharge.9-17 Successful randomized trials have reduced 30-day readmission rates by 20-40%.18 Widespread application of these clinical trial interventions to general practice has also been encouraging. Since 2008, 14 Medicare Quality Improvement Organizations have been funded to focus on care transitions by applying lessons learned from clinical trials. Several have been notably successful in reducing readmissions within 30 days.19 Many of these study interventions involved enhanced clinician involvement and indicate a key role for clinicians in reducing readmissions.9-17

Despite these demonstrated successful interventions, the overall national readmission rate remains high, with a within 30-day readmission following over 15% of discharges. Moreover, we show below that RARRs range from 7% to 25% for eligible clinician groups for 2015-16. Both the high baseline rate and the variability across providers speak to the need for a quality measure to prompt greater care improvement. Given that studies have shown readmissions within 30 days to be related to quality of care, that interventions, including those utilizing clinicians, have been able to reduce 30-day readmission rates for a variety of specific conditions, and that high and variable clinician-level readmission rates indicate opportunity for improvement, we sought to develop measure of all-cause, all-condition 30-day unplanned readmission at the clinician-group level.

References:

1. Frankl SE, Breeling JL, L. G. Preventability of emergent hospital readmission American Journal of Medicine. Jun 1991;90(6):667-674.

2. Corrigan J, Martin J. Identification of factors associated with hospital readmission and development of a predictive model. Health Services Research. Apr 1992;27(1):81-101.

3. Oddone E, Weinberger M, Horner M, et al. Classifying general medicine readmissions. Are they preventable? Veterans Affairs Cooperative Studies in Health Services Group on Primary Care and Hospital Readmissions. Journal of General Internal Medicine. Oct 1996;11(10):597-607.

4. Ashton C, Del Junco DJ, Souchek J, Wray N, Mansyur C. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care. Oct 1997;35(10):1044-1059.

5. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Archives of Internal Medicine. Apr 24, 2000;160(8):1074-1081.

6. Courtney EDJ, Ankrett S, McCollum PT. 28-Day emergency surgical re-admission rates asa clinical indicator of performance. Annals of the Royal College of Surgeons of England. Mar 2003;85(2):75-78.

7. Halfon P, Eggli Y, Pr, et al. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical Care. Nov 2006;44(11):972-981.

8. Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. JAMA. May 5, 2010;303(17):1716-1722.

9. Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M. Comprehensive discharge planning for the hospitalized elderly. A randomized clinical trial. Ann Intern Med. Jun 15 1994;120(12):999-1006.

10. Naylor MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA. 1999;281(7):613-620.

11. Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. Journal of the American College of Cardiology. Jan 2, 2002;39(1):83-89.

12. van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. Journal of General Internal Medicine. Mar 2002;17(3):186-192.

13.Conley RR, Kelly DL, Love RC, McMahon RP. Rehospitalization risk with second-generation and depot antipsychotics. Annals of Clinical Psychiatry. Mar2003;15(1):23-31.

14.Coleman EA, Smith JD, Frank JC, Min S-J, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: The Care Transitions Intervention. Journal of the American Geriatrics Society. Nov 2004;52(11):1817-1825.

15. Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. JAMA. Mar 17, 2004;291(11):1358-1367.

16. Jovicic A, Holroyd-Leduc JM, Straus SE. Effects of self-management intervention on health outcomes of patients with heart failure: a systematic review of randomized controlled trials. BMC Cardiovasc Disord. 2006; 6:43.

17. Garasen H, Windspoll R, Johnsen R. Intermediate care at a community hospital as an alternative to prolonged general hospital care for elderly patients: a randomised controlled trial. BMC Public Health. 2007; 7:69.

18. Leppin AL, Gionfriddo MR, Kessler M, etal. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Intern Med. 2014;174(7):1095-1107.

19. CFMC. CFfMC. Care Transitions QIOSC. 2010; http://www.cfmc.org/caretransitions/.

1b.2. Provide performance scores on the measure as specified (<u>current and over time</u>) at the specified level of analysis*. (<u>This is required for maintenance of endorsement</u>. 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.*

We conducted analyses using a full year of Medicare data from July 2015 through June 2016 (n=6,468,761 included in the final index cohort). We report the of RARRs for eligible clinician groups with at least 25 attributed index admissions. These data provide supportive evidence of performance variation.

Distribution of risk adjusted readmission ratios (RARRs) for eligible clinician groups with at least 25 admissions (dataset: Medicare July 2015-June 2016 Full Sample)

Description//Mean of RARRs in each decile//Eligible Clinician Groups RARR(%)

1//13.1

2//14.1

3//14.6

4//14.9

5//15.2

6//15.5

7//15.8

8//16.2

9//16.7

10//18.0

Description//Distribution of standardized risk ratios//Eligible Clinician Groups RARR(%)

Number of eligible clinician groups//55,593

Mean (SD)// 15.4%(1.4%)

Minimum//7.0

10th percentile//13.8

25th percentile//14.6

50th percentile//15.3

75th percentile//16.2

90th percentile//17.1

Maximum//25.1

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

N/A

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

Distribution of MIPS HWR RARRS by Proportion of Dual Eligible Patients:

Data Source: Medicare FFS claims

Dates of Data: July 2015 through June 2016

Eligible Clinician groups with at least 25 admissions

Characteristic// Eligible Clinician groups with a low proportion (=5.4%) Dual Eligible patients// Eligible Clinician groups with a high proportion (=25.8%) Dual Eligible patients

Number of Measured Entities (Eligible Clinician groups)// 13,935 // 13,914

Number of Patients// 1,569,816 patients in low-proportion Eligible Clinician groups// 1,463,055 in high-proportion Eligible Clinician groups

Maximum// 25.0 // 23.9

90th percentile// 16.8// 17.4

75th percentile// 16.0 // 16.5

Median (50th percentile)// 15.2 // 15.6

25th percentile// 14.5 // 14.8

10th percentile// 13.8 // 14.1

Minimum // 7.0 // 9.7

Distribution of MIPS HWR RARRS by Proportion of Patients with AHRQ SES Index Scores Equal to or Below 42.6:

Data Source: Medicare FFS claims and The American Community Survey (2009-2013) data

Dates of Data: July 2015 through June 2016

Eligible Clinician groups with at least 25 admissions

Characteristic// Eligible Clinician groups with low proportion of patients with AHRQ SES index score equal to or below 42.6 (=7.4%)// Eligible Clinician groups with high proportion of patients with AHRQ SES index score equal to or below 42.6 (=33.3%)

Number of Measured Entities (Eligible Clinician groups)// 13,877 // 13,956

Number of Patients// 1,814,523 patients in Eligible Clinician groups with low proportion of patients with AHRQ SES index score equal to or below 42.6 // 1,492,557 patients in Eligible Clinician groups with high proportion of patients with AHRQ SES index score equal to or below 42.6

Maximum// 23.4 // 24.9

90th percentile// 16.8 // 17.3

75th percentile// 16.0 // 16.4

Median (50th percentile)// 15.2 // 15.5

25th percentile// 14.5 // 14.7

10th percentile// 13.8 // 14.1

Minimum // 7.0 // 9.9

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

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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.)*

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** Del18dHOP5MIPSHWRDataDictionary12172018-637086294768821435.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.

N/A

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

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 readmission within 30-days of a hospital discharge. We define readmission as an inpatient admission for any cause, except for certain planned readmissions, within 30 days from the date of discharge from an eligible index admission.

Additional details are provided in S.5 Numerator Details

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

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

The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index admission, excluding planned readmissions as defined below. The measure outcome is a dichotomous yes or no of whether each discharged patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could

be related to care provided during the intervening planned readmission rather than during the index admission.

Numerator Time Window: The outcome is defined as an unplanned readmission within 30 days of discharge from an index admission.

Planned Readmission Algorithm (Version 4.0)

The Planned Readmission Algorithm is a set of criteria for classifying readmissions as planned among the general Medicare population using Medicare administrative claims data. The algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital.

The Planned Readmission Algorithm has three fundamental principles:

1. A few specific, limited types of care are always considered planned (obstetric delivery,

transplant surgery, maintenance chemotherapy/immunotherapy, rehabilitation);

2. Otherwise, a non-acute readmission for a procedure that is typically scheduled in advance is

considered planned; and

3. Admissions for acute illness or for complications of care are never planned.

The algorithm was developed in 2011 as part of the Hospital-Wide Readmission measure. In 2013, CMS applied the algorithm to its other readmission measures.

The Planned Readmission Algorithm and associated code tables are attached in data field S.2b (Data Dictionary or Code Table).

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

The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from any non-federal, acute care inpatient U.S. hospitals (including territories) with Medicare Part A enrollment for the 12 months prior to admission and Part A enrollment for the 30 days after discharge. These are called 'index admissions'.

Outcome attribution:

There are three eligible clinician groups for attribution: 1) the Primary Inpatient Care Provider, 2) the Discharge Clinician and 3) the Outpatient Primary Care Physician.

1) Primary Inpatient Care Provider: All patient-facing claims for the patient filed during the stay are identified and totaled by clinicians identified on each claim; the admission is attributed to the clinician with the greatest charges billed. The cost of charges billed (as opposed to number of charges) better reflects the appropriate clinician, especially for the surgical specialty cohort. The identified primary inpatient care provider may also be the discharge clinician.

2) Discharge Clinician: Identified by Current Procedural Terminology [CPT®] code 99238 or 99239 within the last three days of admission OR CPTs 99231, 99232, 99233 billed on the last day of admission. If none of these codes found, a Discharge Clinician is not assigned.

3) Outpatient Primary Care Physician: The clinician who provides the greatest number of claims for primary care services during the 12 months prior to the hospital admission date.

Eligible clinician groups are defined by grouping eligible clinicians who use the same Taxpayer Identification Number (TIN). Index admissions are attributed to a clinician group by each of these rules. Though an admission may be attributed to three distinct eligible clinician groups, it will often be the case that two or even all three of the above listed roles for a given patient are filled by clinicians assigned to the same clinician group. In the case of multiple assignments of an admission to the same eligible clinician group, each admission is included only once when measuring the eligible clinician group.

Importantly, this implies that while there are three different rules for attribution, these are not distinguished when measuring clinician group performance. While a clinician group can have admissions attributed to them

in multiple capacities – for instance, a clinician from the same group may be both a Discharge Clinician for some patients and a Primary Inpatient Care Provider for others – all attributed admissions are used to construct a single score for that eligible clinician group. Thus, while we report some results by attribution role, we report measure scores only for "unique eligible clinician groups".

Additional details are provided in S.7 Denominator Details.

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

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

Admissions are eligible for inclusion in the measure if:

1. Patient is 65 or older

Rationale: Younger Medicare patients represent a distinct population with dissimilar characteristics and outcomes.

2. Patient survives index admission

Rationale: Patients who die during the initial admission cannot be readmitted.

3. Patient is not transferred to another hospital

Rationale: In an episode of care in which the patient is transferred between hospitals, responsibility for the readmission is assigned to the final discharging hospital. Therefore, intermediate admissions within a single episode of care are not eligible for inclusion.

4. Patient is continuously enrolled in FFS Medicare Part A for the 12 months prior to the index admission and Part A for 30 days after discharge; FFS Medicare Part B for 12 months prior to index admission.

Rationale: This is necessary to ensure complete data for risk adjustment, attribution, and outcome determination.

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

From the cohort, we exclude admissions if:

1. The patient is discharged against medical advice (AMA)
2. The patient is discharged from a PPS-exempt cancer hospital
3. The patient is admitted primarily for the medical treatment of cancer
4. The patient is admitted primarily for the treatment of psychiatric disease
5. The patient is admitted primarily for "rehabilitation care; fitting of prostheses and adjustment devices" (CCS 254)
6. Admissions without 30 Days of Post-Discharge Enrollment are excluded
7. Admissions cannot be identified in IDR database
8. The admission cannot be attributed to an eligible clinician.

Further exclusion details can be found in S.9 Denominator Exclusion Details

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.)*

From the cohort, we exclude admissions for which:

1. Patients discharged against medical advice (AMA)

Rationale: Clinicians have limited opportunity to implement high quality care

2. Admissions for patients to a PPS-exempt cancer hospital

Rationale: These hospitals care for a unique population of patients that cannot reasonably be compared to the patients admitted to other hospitals.

3. Admissions primarily for medical treatment of cancer are excluded

Rationale: These admissions have a very different mortality and readmission profile compared to the rest of the Medicare population (higher rates of planned readmissions and higher rates of competing mortality), and outcomes for these admissions do not correlate well with outcomes for other admissions. Patients with cancer who are admitted for other diagnoses or for surgical treatment of their cancer remain in the measure.

4. Admissions primarily for psychiatric disease are excluded

Rationale: Patients admitted principally for psychiatric treatment are typically cared for in separate psychiatric centers which are not comparable to acute care hospitals. See Data Dictionary for excluded CCSs.

5. Admissions for "rehabilitation care; fitting of prostheses and adjustment devices" (CCS 254) are excluded

Rationale: These admissions are not typically admitted to an acute care hospital for acute care.

6. Admissions without 30 Days of Post-Discharge Enrollment are excluded

Rationale: The 30-day readmission outcome cannot be assessed in patients who do not maintain enrollment for at least 30 days following discharge.

7. Admissions cannot be identified in IDR database

Rationale: Information from the attribution cannot be applied for patients without data of physician information, which we extracted from IDR database.

8. Patients cannot be attributed to a clinician group.

Rationale: Only patients assigned to eligible clinician groups should be included in the measure.

Note that a readmission within 30-days will also be eligible as an index admission if it meets all other eligibility criteria. This allows our measure to capture repeated admissions for the same patient, whether with the same clinician(s) or not. Since there are few patients with multiple admissions in the same year and in the same specialty cohort, we chose to treat multiple admissions as statistically independent.

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

N/A

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

The index admissions are identified as described above in S.5-S.9.

Specialty Cohorts

The measure uses an algorithm identical to that of the hospital level measure (NQF #1789) to group index admissions into subgroups for risk adjustment. The measure aggregates the ICD-9 and ICD-10 principal diagnosis and all procedure codes of the index admission into clinically coherent groups of conditions and procedures (condition categories or procedure categories) using the AHRQ CCS. There is a total of 285 mutually exclusive AHRQ condition categories, most of which are single, homogenous diseases such as pneumonia or acute myocardial infarction. Some are aggregates of conditions, such as "other bacterial infections." There is a total of 231 mutually exclusive procedure categories. Using these AHRQ CCS procedure and condition categories, the measure assigns each index hospitalization to one of five mutually exclusive specialty cohorts: surgery/gynecology, cardiorespiratory, cardiovascular, neurology, and medicine. The rationale behind this organization is that conditions typically cared for by the same team of clinicians are expected to experience similar added (or reduced) levels of readmission risk.

Step 1. The measure first assigns admissions with qualifying AHRQ procedure categories to the Surgery/Gynecology Cohort. This cohort includes admissions likely cared for by surgical or gynecological teams.

Step 2. The measure then sorts admissions into one of the four remaining specialty cohorts based on the AHRQ diagnosis category of the principal discharge diagnosis:

The Cardiorespiratory Cohort: includes several condition categories with very high readmission rates such as pneumonia, chronic obstructive pulmonary disease, and heart failure. These admissions are combined into a single cohort because they are often clinically indistinguishable and patients are often simultaneously treated for several of these diagnoses.

The Cardiovascular Cohort: includes condition categories such as acute myocardial infarction, that in large hospitals, might be cared for by a separate cardiac or cardiovascular team.

The Neurology Cohort: includes neurologic condition categories such as stroke, that in large hospitals, might be cared for by a separate neurology team.

The Medicine Cohort: includes all non-surgical patients who were not assigned to any of the other cohorts.

The full list of the specific diagnosis and procedure AHRQ CCS categories used to define the specialty cohorts are attached in data field S.2b (Data Dictionary or Code Table).

Risk adjustment

Risk adjustment is done separately for each specialty cohort using a logistic regression model with 30-day readmission as the outcome. Risk adjusters in each model are identical to those used in the specialty cohorts for the hospital level measure (NQF #1789) and include the CCS for the principle diagnosis. The full list of risk adjusters can be found in the Data Dictionary.

Measure Score

Because the same admission may be attributed to more than one unique Eligible Clinician group, we could not apply the method used by the existing hospital-level HWR measure (NQF#1789) to construct risk standardized readmission rates. Instead, we adopted a method that, while requiring an assumption of independence across entities, allowed us to account for correlation within entity. The measure uses instead an approach similar to that used by the Patient Safety and Adverse Events Composite measure (NQF #0531).

Reference the attached Intent to Submit form for the complete response.

Creating Credible Interval Estimates

For purposes of estimating confidence intervals, we used bootstrapping. Because of overlapping assignment of patients, bootstrapping was at the specialty cohort level. Specifically, we select m=1,…,M random samples of discharges with replacement from each specialty cohort. Using the existing attribution, we calculated (1), (2) and (3) above for each provider. The 95% credible interval estimate of the RARRj for each provider was used as the estimated 95% confidence interval.

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims, 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

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)*

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

MIPSHWRTestingForm10.31.19_v1.0.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.

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.

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.

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

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

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

Measure Title: Hospital-Wide, 30-Day, All-Cause Unplanned Readmission (HWR) Rate for the Merit-Based Incentive Payment System (MIPS) Eligible Clinician Groups

Date of Submission: 12/17/2018

Type of Measure:

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

Instructions

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

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

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

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

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

AND

If patient preference (e.g., informed 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 nonresponders) and how the specified handling of missing data minimizes bias.

Notes

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

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

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

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

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

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

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

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*).

Testing was performed using Medicare Part A and B claims data, as well as the Medicare Enrollment Database (EDB). See section 1.7 for details.

1.3. What are the dates of the data used in testing? Click here to enter date range

The dates used varied by testing type; see section 1.7 for details.

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

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

1.5. How many and which <u>measured entities</u> 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)

Eligible Clinician Groups participating in the Merit-Based Incentive Payment System (MIPS) are the measured entities; Medicare Fee-for-Service (FFS) beneficiaries aged 65 years and older are included. Eligible Clinician Groups are identified by aggregating NPI-TIN pairs with a common TIN. The number of measured entities (Eligible Clinician Groups) varies by testing type; see Section 1.7 for details.

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

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

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

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

<u>Measure Development and Testing</u>

For measure development and testing, we used two-years (July 2015 – June 2017) of Medicare administrative claims data. The dataset contained Medicare inpatient and outpatient claims data and Medicare enrollment information. We randomly split the first year of data into two equal samples: **Dataset A1** (July 2015 – June 2016) was used as the "development sample" and **Dataset A2** (July 2015 – June 2016) was used as the "validation sample." We used the second year of the data in addition to an extra year of data, **Dataset B,** as a temporal validation sample set (July 2016 – June 2017).

Dataset A1 (Development sample; July 2015 – June 2016):

Number of admissions = 3,234,836

Number of clinician groups = 117,788

Patient Descriptive Characteristics: mean age = 78.21 years; % female = 56.29

Dataset A2 (Validation sample; July 2015 – June 2016):

Number of admissions = 3,233,925

Number of clinician groups = 117,693

Patient Descriptive Characteristics: mean age = 78.21 years; % female = 56.33

Dataset B (Temporal validation sample; July 2016 – June 2017):

Number of admissions = 6,411,508

Number of clinician groups = 124,311

Patient Descriptive Characteristics: mean age = 78.19 years; % female = 56.05

For reliability testing (Section 2a2)

The measure development and validation samples were used to assess measure score reliability and for statistical model testing.

Datasets A1, A2, Dataset B was used for data elements reliability, patient level model reliability

For testing of measure exclusions (Section 2b2)

We used one year of data from June 2015 -July 2016 to test for measure exclusions (Dataset A1/A2 Combined)

For testing of measure risk adjustment (Section 2b3)

We used two years of data from July 2015 - June 2017 (Dataset A1/A2/B Combined)

For testing to identify meaningful differences in performance (Section 2b4)

We used one year of data from July 2015 – June 2016 (Dataset A1/A2 combined).

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

We selected the social risk factors to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various social risk factors to worse health status, higher mortality over a lifetime, and hospital outcomes such as readmission and complications.[1-9] The causal pathways for socioeconomic status (SES) variable selection are described below in Section 2b4.3.

The SES variables used for analysis were:
• Dual-eligible status: Dual-eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data is obtained from the Master Beneficiary Summary File (MBSF). The MBSF is an annually created file that contains enrollment information for all Medicare beneficiaries including dual-eligible status. Year 2014-2017 were used.
• Agency for Healthcare Research and Quality (AHRQ)-validated SES index score: AHRQ SES index score is based on a beneficiary 9-digit zip code level of residence and incorporates 7 census variables found in The American Community Survey (2009-2013).

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.

In selecting variables, our intent was to be responsive to the NQF guidelines for measure developers. Our approach has been to examine all patient-level indicators of both SES that are reliably available for all Medicare beneficiaries, are linkable to claims data, and have established validity.

We recognize that Medicare-Medicaid dual-eligibility has limitations as a proxy for patients' income or assets because it does not provide a range of results and is only a dichotomous indicator. However, the dual-eligibility threshold for over 65-year-old Medicare patients is a meaningful marker for SDS, as it accounts for both income and assets and is consistently applied across states for the older population. For dual-eligibility, there is a body of literature demonstrating differential health care and health outcomes among beneficiaries indicating that these variables, while not ideal, also allow us to examine some of the pathways of interest.[10]

Finally, we selected the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas.[11] Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey(ACS) data. A census block group is a geographical unit used by the U.S. Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2009-2013 ACS data and mapped patients' ZIP codes via vendor software to the AHRQ SES Index at the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). This provides a better marker of low SES neighborhoods in high expense geographic areas. We then calculated an AHRQ SES Index score for census block groups that can be linked to 9-digit ZIP codes. In the THA/TKA measure cohort, we were able to assign an AHRQ SES Index score to 99.5% of patient admissions. 88.7% of patient admissions had calculated AHRQ SES Index scores linked to their 9-digit ZIP codes. 10.8% of patient admissions had only valid 5-digit ZIP codes; we utilized the data for the median 9-digit ZIP code within that 5-digit ZIP code.

References:
1. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health affairs

(Project Hope). 2002; 21(2):60-76.

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

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

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

5. Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health affairs (Project Hope). 2014; 33(5):778-785.

6. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. Jan 23 2013; 309(4):342-3.

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

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

9. van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. European journal of epidemiology. Aug 2012; 27(8):605-613.

10. Assistant Secretary for Planning and Evaluation (ASPE). Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed June 21, 2018.

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

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: Clinician Group-Level Reliability

We estimated the clinician group signal-to-noise reliability for each of the five specialty cohorts (that are combined to produce the overall measure result). This is "unit" reliability, that is, the reliability with which individual units (here, clinician groups) are measured. Because signal-to-noise reliability is based on model parameters, it is only meaningful to calculate it at the level of the specialty cohort; however, according to Rudner[1], the reliability of an aggregated scale is bounded below by the reliability of the least reliable component, and will generally be greater than the most reliable component of the component scales are positively correlated. For each cohort, we use the formula presented by Adams et al[2], to calculate provider-level signal-to-noise reliability. For each measured entity (eligible clinician group) we calculated the ratio of $\tau^2/(\tau^2 + \sigma^2)$, using the value τ^2 defined in section S.14 of the Intent to Submit form and estimating σ^2 using on bootstrapping. We summarized the distribution of these values for each of the 5 specialty cohorts and all groups with at least 25 attributed index admissions in the cohort. Because the reliability of any one entity's measure score will generally increase with the number of index admissions attributed, groups with higher volume will tend to have more reliable scores, while those with lower volume will tend to have less reliable scores. For this reason, we also report the mean and range for different volume subgroups for each cohort.

References

1. Rudner, L (2001) 'Informed test component weighting', Educational Measurement: Issues and Practice, 20, pp. 16-19

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

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*)

Mean Signal to Noise Ratio at clinical group level

Variable	Mean	StdDev	Min	P5	P10	Q1	Median	Q3	P90	P95	Max
Cardiorespiratory	0.55	0.18	0.22	0.33	0.35	0.41	0.51	0.66	0.84	0.90	0.99
CV	0.57	0.18	0.25	0.34	0.36	0.42	0.53	0.71	0.86	0.91	0.99
Medicine	0.47	0.19	0.15	0.23	0.26	0.31	0.42	0.59	0.77	0.87	1.00
Neurology	0.65	0.16	0.35	0.44	0.46	0.51	0.63	0.79	0.89	0.93	0.99
Surgical	0.45	0.20	0.08	0.21	0.23	0.29	0.40	0.57	0.78	0.87	0.99

Mean Signal to Noise Ratio at clinical group level at different case volume thresholds

Clinician Group Case Volume	Cardio-respiratory Mean (Range)	Cardiovascular Mean (Range)	Medicine Mean (Range)	Neurology Mean (Range)	Surgical Mean (Range)
25+	0.56 (0.22 - 0.99)	0.57 (0.25 - 0.99)	0.47 (0.15 - 0.99)	0.65 (0.36 - 0.99)	0.45 (0.08 - 0.99)
[25,100)	0.41 (0.22 - 0.63)	0.42 (0.25 - 0.66)	0.31 (0.15 - 0.61)	0.52 (0.36 - 0.73)	0.31 (0.08 - 0.59)
[50, 100)	0.58 (0.39 - 0.75)	0.59(0.43 - 0.79)	0.47 (0.25 - 0.71)	0.68 (0.54 - 0.83)	0.45 (0.15 - 0.77)
[100,150)	0.71 (0.59 - 0.81)	0.72 (0.62 - 0.85)	0.62 (0.44 - 0.76)	0.79 (0.72 - 0.88)	0.58 (0.28 - 0.81)
[150,200)	0.78 (0.68 - 0.84)	0.78 (0.70 - 0.88)	0.69 (0.55 - 0.81)	0.84 (0.81 - 0.91)	0.66 (0.38 - 0.85)
200+	0.89 (0.77 - 0.99)	0.88 (0.74 - 0.99)	0.85 (0.64 - 0.99)	0.92 (0.85 - 0.99)	0.82 (0.43 - 0.99)

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

Measure Score Reliability Results

The mean signal-to-noise reliability scores of the five cohorts ranged from 0.45 to 0.65 for clinician groups, calculated with 1 year of data and providers with at least 25 patients in the cohorts, are considered "substantial" for clinician groups.

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

< 0 –Less than chance agreement;
0 –0.2 Slight agreement;
0.21 –0.39 Fair agreement;
0.4 –0.59 Moderate agreement;
0.6 –0.79 Substantial agreement;
0.8 – 0.99 Almost Perfect agreement; and
1.0 Perfect agreement

Reference

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

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Measure validity is demonstrated through empirical validity testing, by systematic assessment of measure face validity via a technical expert panel (TEP) of national experts and stakeholder organizations, and through use of established measure development guidelines.[1-6]

Empirical Validity Testing

To provide additional validation of the measure, we assessed how well it correlated with hospital quality. Since the outcome depends on hospital processes, including coordination of care during the stay and during transition from the hospital, we anticipate the readmission rate for a provider would be consistent with the quality of the hospital where most of their attributed patients are discharged. As measures of hospital quality, we used the CMS Hospital Overall Star Ratings and Hospital Star Ratings readmission domain scores. To assess consistency, we plotted the distribution of the measure score, Risk Adjusted Readmission Rate (RARR) over: a) Overall Star Rating (1-5) and b) quintiles of the Star Rating Readmission domain score for Eligible Clinician Groups with at least 25 patients attributed.

Validity as Assessed by External Groups and TEP

Throughout the measure development process, we obtained expert and stakeholder input through holding regular discussions with external clinical consultants, consulting our national TEP, and holding a 30-day public comment period.

CORE clinicians, as well as several clinical experts, met regularly to discuss all aspects of measure development, including the cohort, outcome definition, risk adjustment and attribution rules.

In addition to the clinical consultations and in alignment with CMS 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 three 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 attribution based on TEP feedback on the measure.
TEP members were asked to assess the face validity of the final measure specification by confidentially responding to two questions:

The risk-standardized readmission rates obtained from the MIPS HWR measure as specified:

1. Are valid and useful measures of MIPS Eligible Clinicians and MIPS Eligible Clinician Group quality of care.
2. Will provide MIPS Eligible Clinicians and MIPS Eligible Clinician Groups with information that can be used to improve their quality of care.

TEP members were asked to report their agreement with each statement on a 6-point scale, representing a range from “strongly disagree” to “strongly agree.”

We also solicited input from the CMS Quality Payment Program (QPP) Clinical Champions specifically around the attribution approach.

Finally, following completion of the preliminary measure, we solicited public comment on the measure through the CMS site:

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 outcomes measurement set forth in NQF guidance for outcome measures[7], CMS Measure Management System (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]

References

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*)

***RARR vs Star Rating (25+)**

Star Rating	Eligible clinicians Groups				
	1	2	3	4	5
Mean	15.8	15.6	15.4	15.2	15.0
Std Dev	1.4	1.4	1.3	1.3	1.4
Minimum	10.3	9.7	7.0	8.8	8.0
5th Percentile	13.8	13.6	13.4	13.1	12.7
10th Percentile	14.2	14.0	13.9	13.6	13.3
25th Percentile	14.9	14.7	14.6	14.4	14.2

Median	15.7	15.5	15.3	15.2	15.0
75th Percentile	16.6	16.4	16.1	16.0	15.8
90th Percentile	17.6	17.3	17.0	16.8	16.6
95th Percentile	18.2	17.9	17.6	17.4	17.1
Maximum	25.1	24.9	23.3	25.0	23.2

***RARR vs Readmission Score from Star Rating (25+)**

Quintile	Eligible clinicians Groups				
	0% ~20% (Worst)	**20%~40%**	**40%~60%**	**60%~80%**	**80% ~100% (Best)**
Mean	15.9	15.6	15.4	15.2	15.0
Std D.	1.4	1.3	1.3	1.3	1.3
Minimum	9.7	8.8	8.0	9.2	7.0
5th Percentile	13.8	13.6	13.4	13.2	12.8
10th Percentile	14.3	14.1	13.9	13.7	13.4
25th Percentile	15.0	14.8	14.6	14.4	14.2
Median	15.8	15.5	15.3	15.2	15.0
75th Percentile	16.7	16.4	16.1	15.9	15.7
90th Percentile	17.7	17.3	16.9	16.7	16.5
95th Percentile	18.3	17.9	17.5	17.3	17.0
Maximum	25.1	23.1	25.0	24.9	21.5

Empirical Validity Testing

The results of assessing the distribution of Eligible Clinician Group RARRs over Star Ratings and readmission score quintiles are shown in Figures 1 and 2.

Figure 1. Distribution of Eligible Clinician Group measure scores over Star Ratings

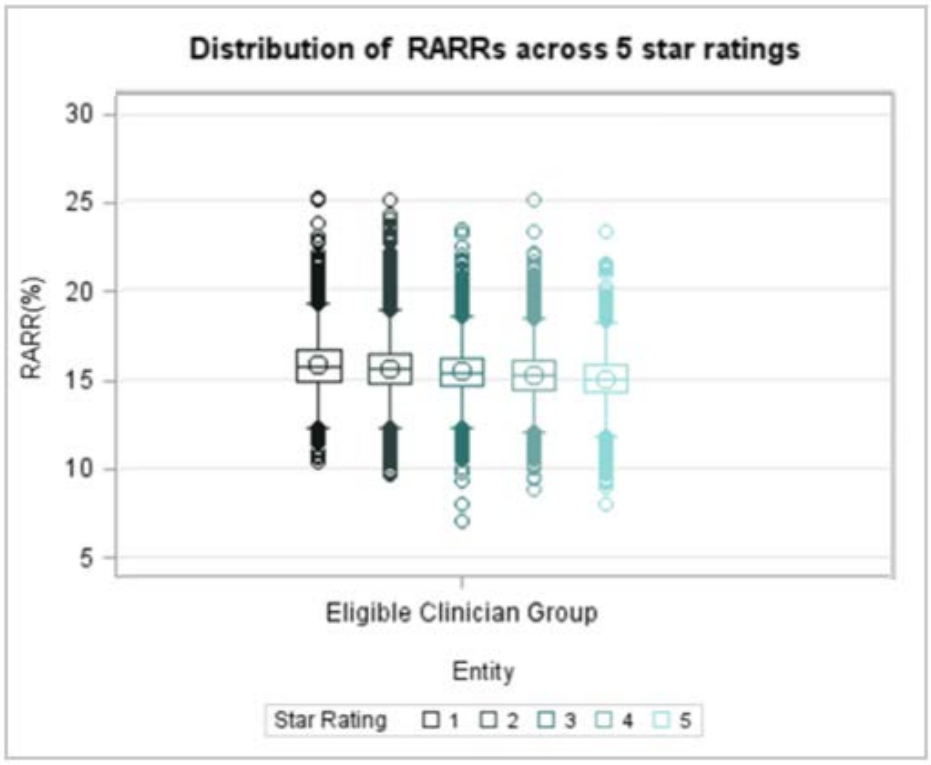
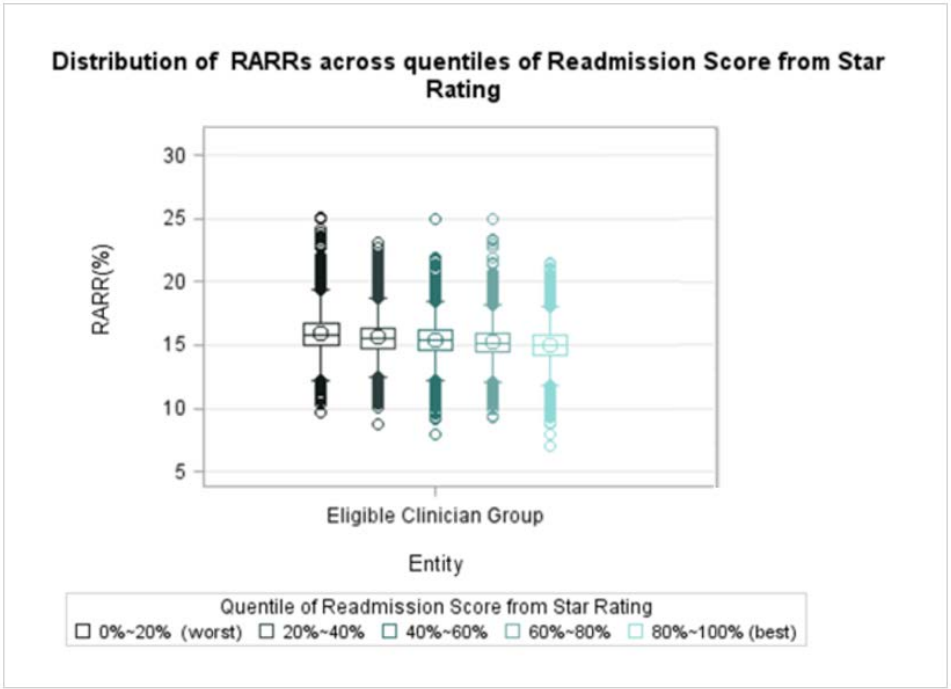


Figure 2. Distribution of Eligible Clinician Group measure scores over quintiles of Star Ratings readmission score



Face Validity as Determined by TEP

Validity was assessed by the TEP. Of 19 TEP members asked to complete a survey regarding validity and usability of the measure, 17 responded.

The majority of respondents, 12/17 or 70%, agreed that the HWR measure scores were valid and useful, and the same proportion agreed that the measure would provide information that could be used to improve the quality of care.

The TEP supported attribution to multiple clinician groups. The TEP considered the fact that readmission is a multi-factorial outcome and that multiple clinicians may play important roles in providing appropriate care, making good decisions and practical recommendations, and intentionally planning care transitions. These actions, while only part of the overall success of a care transition, are necessary to promote the best possible outcome. In this context, most TEP members agreed that patient outcomes should not be attributed to a single clinician but should rather be attributed to multiple clinicians responsible for the patient during and after the inpatient stay. Some TEP members noted the multiple attribution algorithm addressed the current reality of patient care, as opposed to being optimistic in how care should be coordinated. Some TEP members also felt that clinicians in the hospital were best able to drive system changes. TEP members were also supportive of attributing to the POC as they felt it was in part the responsibility of the outpatient provider to ensure the patient does not return to the hospital with remnant issues from their inpatient stay.

Among those who disagreed, the primary concern was that factors which led to increased risk of readmission were beyond the control of any single eligible clinician or clinician group. This concern drove the adoption of 'multiple' attribution, in which no single eligible clinician is solely responsible for a readmission outcome; this attribution approach also has the potential to incentivize collaboration within the hospital and across the care system, further aligning the measure with the attribution.

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

Empirical Validity Testing

The results of the external validity testing as shown in the figures in 2.b1.3 indicate that Eligible Clinician Group RARRs go down with increasing overall hospital quality Star Rating and with increasing quintile of the Star Rating readmission quality score. These provide meaningful external validation of the current measure.

Validity as Assessed by External Groups and by TEP

The results demonstrate TEP agreement with overall face validity of the measure as specified.

Measure validity is also ensured through the processes employed during development, including regular expert and clinical input.

The attribution rules adopted by this measure were also reviewed and supported by the TEP members as well as several clinical expert workgroup members.

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 (Dataset A1/A2 Combined). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

Given that we used the hospital cohort as the initial cohort for attributing to clinician groups, we report the distribution of each exclusion at hospital level. We anticipate that any bias in the clinician group level would be similar to that at the hospital level.

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

In DatasetA1/A2 Combined:

Exclusion	N	%	Distribution across hospitals (N=4,723): Minimum, 25th percentile, 50th percentile, 75th percentile, maximum
Admitted to PPS-Exempt Cancer Hospitals	21,262	0.30	(0.00, 0.00, 0.00, 0.00, 100.00)
Without 30 Days of Post-Discharge Enrollment	33,813	0.48	(0.00, 0.00, 0.36, 0.65, 20.00)
Discharged against medical advice(AMA)	28,708	0.41	(0.00, 0.00, 0.24, 0.53, 20.00)
Admitted for Primary Psychiatric Diagnosis	20,101	0.29	(0.00, 0.00, 0.08, 0.26, 85.19)
Admitted for Rehabilitation	1,554	0.02	(0.00, 0.00, 0.00, 0.00, 100.00)
Admitted for Medical Treatment of Cancer	148,575	2.11	(0.00, 0.40, 1.20, 1.88, 67.91)
Without data matched in IDR database	172,160	2.45	(0.00, 0.38, 0.91, 1.79, 100.00)
Without attribution assignment	149,831	2.13	(0.00, 0.47, 1.27, 2.43, 88.89)

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

Exclusion applied to the HWR measure cohort

1. Patients admitted to Inpatient Prospective Payment System (IPPS)-exempt cancer hospitals account for 0.30% of all index admissions excluded from the initial cohort. Admissions for treatment of cancer are associated with a very different mortality and readmission risk compared with admissions to other IPPS hospitals for treatment of other diseases. Additionally, outcomes for these admissions do not correlate well with outcomes for other types of admissions. (Patients with cancer who are admitted for other diagnoses or for surgical treatment of their cancer remain in the measure).

2. Patients without at least 30 days post-discharge enrollment in FFS Medicare following discharge account for 0.48% of all index admissions excluded from the initial cohort. This exclusion is needed since the 30-day readmission outcome cannot be assessed in patients who do not maintain enrollment for at least 30 days following discharge.

3. Patients discharged against medical advice (AMA) account for 0.41% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to adequately deliver full care and prepare the patient for discharge.

4. Patients admitted for primary psychiatric diagnoses account for 0.29% of all index admissions excluded from the initial cohort. This exclusion is needed because these patients are typically cared for in separate psychiatric or rehabilitation centers which are not comparable to acute care hospitals.

5. Patients admitted for rehabilitation account for 0.02% of all index admissions excluded from the initial cohort. This exclusion is needed because patients admitted for rehabilitation are not admitted for treatment of acute illness and the care provided in rehabilitation centers is not comparable to care provided in acute care hospitals.

6. Patients admitted for medical treatment of cancer account for 2.11% of all index admissions excluded from the initial cohort. Admissions for treatment of cancer are associated with a very different mortality and readmission risk compared with admissions to other IPPS hospitals for treatment of other diseases. Additionally, outcomes for these admissions do not correlate well with outcomes for other types of admissions. (Patients with cancer who are admitted for other diagnoses or for surgical treatment of their cancer remain in the measure).

7. Patients whose index admissions cannot be matched with data in IDR database for 2.45% of all index admission excluded from the initial cohort, where the Part B carrier lines from IDR database are used to identify the eligible clinician groups. Without the ability to attribute the index admission to an eligible clinician group, we cannot assign responsibility for the measure outcome.

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** 32 + variable number of condition category (CC) **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

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

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

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

N/A. This measure will be 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 based on a conceptual model that 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" (Krumholz HM, et al., 2006). The measure estimates clinician-level 30-day all-cause risk-adjusted readmission rates using reliability-adjusted observed to expected ratios. In brief, the approach first models data at the patient level to construct clinician group level observed to expected ratios, and then adjusts these ratios to account for variance in patient outcomes within and between hospitals (Hardy and Thompson, 1996).

As with the hospital measure, admissions are assigned to one of five mutually exclusive specialty cohort groups consisting of related conditions or procedures. At the patient level, the measure models within each cohort the log-odds of hospital readmission within 30 days of discharge using age and selected clinical covariates representing case mix and service mix. This patient level model is used to construct an expected probability of a readmission for each patient; for each clinician group, the total observed readmissions and total probabilities of readmission are calculated by summing over all attributed patients. The ratio of these two totals, observed/expected (O/E), is then adjusted across all clinician groups using a method that incorporates both the provider specific estimate and the overall mean O/E scores, to create a standardized ratio (SR).

The specialty cohort SRs are then pooled for each clinician group using a volume-weighted geometric mean to create a hospital-wide composite SR. The composite SR is multiplied by the national observed readmission rate to produce the RARR.

Data Source
The HWR risk-adjustment models use only inpatient claims data (history and current) in order to make it feasible to implement with Medicare data, and to make it applicable to all-payer data, which are typically restricted to inpatient claims.

The HWR measure uses CMS-CCs (Horwitz L, Partovian C, Lin Z, et al. 2012), the grouper used in previous CMS risk-standardized outcomes measures, to group ICD-9-CM codes into comorbid risk adjustment variables, since four CMS condition-specific claims-based readmission models that use this grouper to define variables for risk adjustment have been validated against models that use chart-abstracted data for risk

adjustment (Pope G, et al., 2000, Keenan PS, Normand SL, Lin Z, et al., 2008, Krumholz HM, Lin Z, Drye EE, et al. 2011).

Approach to Variable Selection:
To harmonize with the existing hospital level measure (NQF #1789) the same risk factors were adopted. These were originally selected from a pool comprised of 30 variables drawn from previous readmission measures (acute myocardial infarction, heart failure, pneumonia, hip and knee arthroplasty, and stroke) and 11 additional CMS-CCs that were determined on a clinical basis to be relevant to an all-condition measure. Using data from the index admission and any admission in the prior 12 months, standard logistic regression models were run for every discharge condition category with the full set of candidate risk adjustment variables. The final set of comorbid risk variables were selected for the hospital measure (NQF #1789) based on the following principles:
• Exclude risk variables that were statistically significant for very few condition categories, given that they would not contribute much to the overall models
• Exclude risk variables that behaved in clinically incoherent ways
• Exclude risk variables that were predominantly protective when it was felt this protective effect was not clinically reasonable but more likely reflected coding factors
• Risk variables that were clinically coherent and carried similar risks across condition categories were grouped. For example, we combined coronary artery disease (CCs 83-84) with cerebrovascular disease (CCs 98, 99, and 103).
• Risk variables that had been combined in previous CMS publicly reported measures were examined, and in one instance separated.

Complications occurring during hospitalization are not comorbid illnesses, may reflect hospital quality of care, and therefore should not be used for risk adjustment. Hence, conditions that may represent adverse outcomes due to care received during the index hospital stay are not included in the risk-adjusted model.

Service mix adjustment:
The measure includes many different discharge condition categories that differ in their baseline readmission risks. In addition, hospitals differ in their relative distribution of these condition categories (service mix). To adjust for service mix, the measure uses an indicator variable for the discharge condition category in addition to risk variables for comorbid conditions. The models include a condition-specific indicator for all condition categories with sufficient volume (defined as those with more than 1,000 admissions nationally in a given year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model.

Socioeconomic Status (SES) Factors
SES factors for examination were based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SES may influence 30-day readmission.

Our conceptualization of the pathways by which patient SES affects 30-day readmission is informed by the literature.

SES and Readmission
As part of the original hospital measure, a literature search was performed on the relationship between SES variables and hospital 30-day, hospital-wide, all-cause, unplanned readmission following hospitalization. The search had the following exclusion criteria: international studies, articles published more than 10 years earlier, articles without primary data, articles using Veterans Affairs (VA) databases as the primary data source, and articles not explicitly focused on SES and readmission across multiple conditions. One hundred and sixty-nine articles were initially reviewed, and one hundred and fifty-five studies were excluded from full-text review based on the above criteria. Studies indicate that SES variables were associated with

increased risk of readmission across multiple major illnesses and conditions (Aseltine RH, et al., 2015; Mitchell SE, et al., 2012; Odonkor CA, et al., 2015; Herrin J, et al., 2015; Gu Q, et al., 2014, Kim H, et al., 2010; Kangovi S, et al., 2012; Iloabuchi TC, 2014; Beck AF, et al., 2012; Arbaje AI, et al., 2008; Hu J, 2014; Nagasako EM, et al., 2014; Joynt, KE, et al., 2013), though there may not be a significant effect on hospital-level profiling (Blum AB, et al., 2014).

SES Variable Selection

Although some recent literature evaluates the relationship between patient SES and the readmission outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with readmission. The SES factors that have been examined in the readmission literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients, and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Eapen ZJ, et al., 2015; Hu J, et al., 2014). Neighborhood/community-level variables use information from sources such as the American Community Survey (ACS) as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the AHRQ-validated SES index score (Blum AB, et al., 2014). Hospital-level variables measure attributes of the hospital which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital (Gilman M, et al., 2014; Joynt KE and Jha AK, 2013).

The conceptual relationship, or potential causal pathways by which these possible SES risk factors influence the risk of readmission following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider. While this measure attributes readmissions to clinician groups, it still captures quality care provided by clinicians within the hospital system. Thus, the conceptual framework used to understand how social risk factors play a role in how hospitals influence readmissions can also be used to understand how social risk factors play a role in how clinician groups at the hospital can influence readmissions.

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

2. Use of low-quality hospitals. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus patients with low income are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission following hospitalization (Jha AK, et al., 2011; Reames BN, et al., 2014). Similarly, African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner J, et al., 2005).

3. Differential care within a hospital. The third major pathway by which SES factors may contribute to readmission risk is that patients may not receive equivalent care within a facility. Alternatively, patients with SES risk factors such as lower education may require differentiated care – e.g. provision of lower literacy information – that they do not receive.

4. Influence of SES on readmission risk outside of hospital quality and health status. Some SES risk factors, such as income or wealth, may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make

appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are complex to distinguish analytically. They also have different implications on the decision to risk adjust or not. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in Section 1.8, the following SES variables were considered:
• Dual-eligible status
• AHRQ SES index

We first considered adjustment for clinical conditions and then examined additional risk imparted by social risk factors after the potential for greater disease burden is included in the risk model. We believe that this is consistent with NQF current guidance and is appropriate given the evidence that people who experience greater social risk are more likely to have more disease burden compared with those who do not.

We assessed the relationship between these variables and the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of this variables improved model performance or changed hospital results. See 2.3b.4
We assessed the relationship between the SES variables with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

References:
Arbaje AI, Wolff JL, Yu Q, Powe NR, Anderson GF, Boult C. Postdischarge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. The Gerontologist. 2008;48(4):495-504
Aseltine RH, Jr., Yan J, Gruss CB, Wagner C, Katz M. Connecticut Hospital Readmissions Related to Chest Pain and Heart Failure: Differences by Race, Ethnicity, and Payer. Connecticut medicine. 2015;79(2):69-76.
Beck AF, Simmons JM, Huang B, Kahn RS. Geomedicine: area-based socioeconomic measures for assessing risk of hospital reutilization among children admitted for asthma. American journal of public health. 2012;102(12):2308-2314.
Blum AB, NN Egorova, E. A. Sosunov, A. C. Gelijns, E. DuPree, A. J. Moskowitz, A. D. Federman, D. D. Ascheim and S. Keyhani. "Impact of Socioeconomic Status Measures on Hospital Profiling in New York City." Circ Cardiovasc Qual Outcomes 7, no. 3 (2014): 391-7.
Eapen ZJ, McCoy LA, Fonarow GC, Yancy CW, Miranda ML, Peterson ED, Califf RM, Hernandez AF. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.
Gilman M, Adams EK, Hockenberry JM, Wilson IB, Milstein AS, Becker ER. California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Aff (Millwood). Aug 2014; 33(8):1314-22.
Gu Q, Koenig L, Faerberg J, Steinberg CR, Vaz C, Wheatley MP. The Medicare Hospital Readmissions Reduction Program: potential unintended consequences for hospitals serving vulnerable populations. Health services research. 2014;49(3):818-837.
Hardy RJ, Thompson SG. A Likelihood Approach to Meta-analysis with random effects. Statistics in medicine. 1996;15:619-629.
Herrin J, St Andre J, Kenward K, Joshi MS, Audet AM, Hines SC. Community factors and hospital readmission rates. Health services research. 2015;50(1):20-39.
Horwitz L, Partovian C, Lin Z, et al. Hospital-Wide All-Cause Unplanned Readmission Measure: Final Technical Report. 2012;

https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228772504318
Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health affairs (Project Hope). 2014; 33(5):778-785.
Iloabuchi TC, Mi D, Tu W, Counsell SR. Risk factors for early hospital readmission in low-income elderly adults. Journal of the American Geriatrics Society. 2014;62(3):489-494.
Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and Medicaid patients. Health Affairs 2011; 30:1904-11.
Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. Jan 23 2013; 309(4):342-3.
Joynt, K. E., E. J. Orav and A. K. Jha. "Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." JAMA 305, no. 7 (2011): 675-81.
Kangovi S, Grande D, Meehan P, Mitra N, Shannon R, Long JA. Perceptions of readmitted patients on the transition from hospital to home. Journal of hospital medicine. 2012;7(9):709-712.
Keenan PS, Normand SL, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circulation 2008;1(1):29-37.
Kim H, Ross JS, Melkus GD, Zhao Z, Boockvar K. Scheduled and unscheduled hospital readmissions among patients with diabetes. The American journal of managed care. 2010;16(10):760-767.
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.
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 2011;4(2):243-52.
Mitchell SE, Sadikova E, Jack BW, Paasche-Orlow MK. Health literacy and 30-day postdischarge hospital utilization. Journal of health communication. 2012;17 Suppl 3:325-338.Nagasako EM, Reidhead M, Waterman B, Dunagan WC. Adding socioeconomic data to hospital readmissions calculations may produce more useful results. Health affairs (Project Hope). 2014;33(5):786-791
Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.
Odonkor CA, Hurst PV, Kondo N, Makary MA, Pronovost PJ. Beyond the Hospital Gates: Elucidating the Interactive Association of Social Support, Depressive Symptoms, and Physical Function with 30-Day Readmissions. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists.* 2015;94(7):555-567
Pope, G., et al., Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review, 2000. 21(3):26.
Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.
Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

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?

N/A

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.**

Variation in Prevalence of the Factor across Measured Entities

The median percentage of Medicaid patients is 12.89% (interquartile range [IQR] 5.41% to 25.40%). The median percentage of low SES AHRQ indicator patients is 17.65% (IQR 7.41% to 33.33%).

Empirical association with the outcome (bivariate)

The patient-level observed hospital wide readmission rate is higher for dual-eligible patients, 19.51%, compared with 14.63% for all other patients. Similarly, the readmission rate for patients in the lowest SES quartile by AHRQ Index was 17.42% compared with 14.90% for all other patients.

Incremental Effect of SES Variables in a Multivariable Model

We then examined the strength and significance of the SDS variables in each of the five specialties cohort multivariable models. When we include any of these variables in a multivariate model that includes all the claims-based clinical variables the effect size of each of these variables is modest, ranging from 1.05 to 1.16; see table below.

	Dual-Eligible	**Low SES**
Cohort	**Odds Ratio**	**Odds Ratio**
CARDIORESPIRATORY	1.12	1.10
CV	1.13	1.09
MEDICINE	1.07	1.05
NEUROLOGY	1.10	1.07
SURGICAL	1.16	1.08

We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model

Cohort	**Dual-Eligible**	**Low SES**	**No SES factor**
CARDIORESPIRATORY	0.641	0.641	0.641
CV	0.664	0.664	0.664
MEDICINE	0.648	0.648	0.648
NEUROLOGY	0.632	0.632	0.631
SURGICAL	0.701	0.701	0.700

Furthermore, we find that the addition of any of these variables into the model has little to no effect on eligible clinician group performance. We examined the change in eligible clinician group RARRs with the addition of any of these variables.

The median absolute change in Eligible Clinician Group RARRs when adding a dual-eligible indicator is -0.02% (interquartile range [IQR] -0.05% to 0.02%, minimum -0.45%, maximum 1.74%) with a correlation coefficient between RARRs for each Eligible Clinician Group with and without Medicaid added of 0.9982. The median absolute change in Eligible Clinician Group's RARRs when adding a low SES AHRQ indicator is -0.01% (IQR -0.03% to 0.02%, minimum -2.24%, maximum 2.45%) with a correlation coefficient between RARRs for each Eligible Clinician Group with and without low SES added of 0.9972.

Summary

We found wide variation in the distribution of the social risk factors we examined, and we found that both had some association with readmission risk. However, adjustment for these factors did not improve model performance nor have an appreciable impact on entity RARRs, suggesting that existing clinical risk factors capture much of the risk related to low SES. Therefore, we did not include SES factors in our final risk model.

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 computed three summary statistics and model performance across the development and validation samples (Dataset A1, A2, and B). We conducted:

Discrimination statistics

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

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

Calibration statistics

3. The calibration value of close to zero at one end and close to 1 on the other end indicates good calibration of the model. See section 2a2.3. for detailed calibration statistics results.

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

We tested the performance of the model for **Dataset A1, A2, and B** described in section 1.7.

References

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

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

If stratified, skip to 2b3.9

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

Results for the development cohort (Dataset A1, A2 and B)

C-statistics

Cohort	2015-2016 Development Sample	2015-2016 Validation Sample	2016-2017 Temporal Validation Sample
Cardiorespiratory	0.64	0.64	0.64
Cardiovascular	0.66	0.66	0.66
Medicine	0.65	0.65	0.65
Neurology	0.63	0.63	0.63
Surgery/Gynecology	0.70	0.70	0.71

Predictive ability (lowest decile %, highest decile %):

Cohort	2015-2016 Development Sample	2015-2016 Validation Sample	2016-2017 Temporal Validation Sample
Cardiorespiratory	9.76 - 35.94	9.78 - 35.68	9.43 - 35.3
Cardiovascular	6.86 - 31.81	6.8 - 31.72	6.74 - 31.79
Medicine	8.48 - 33.69	8.44 - 33.73	8.66 - 34.13

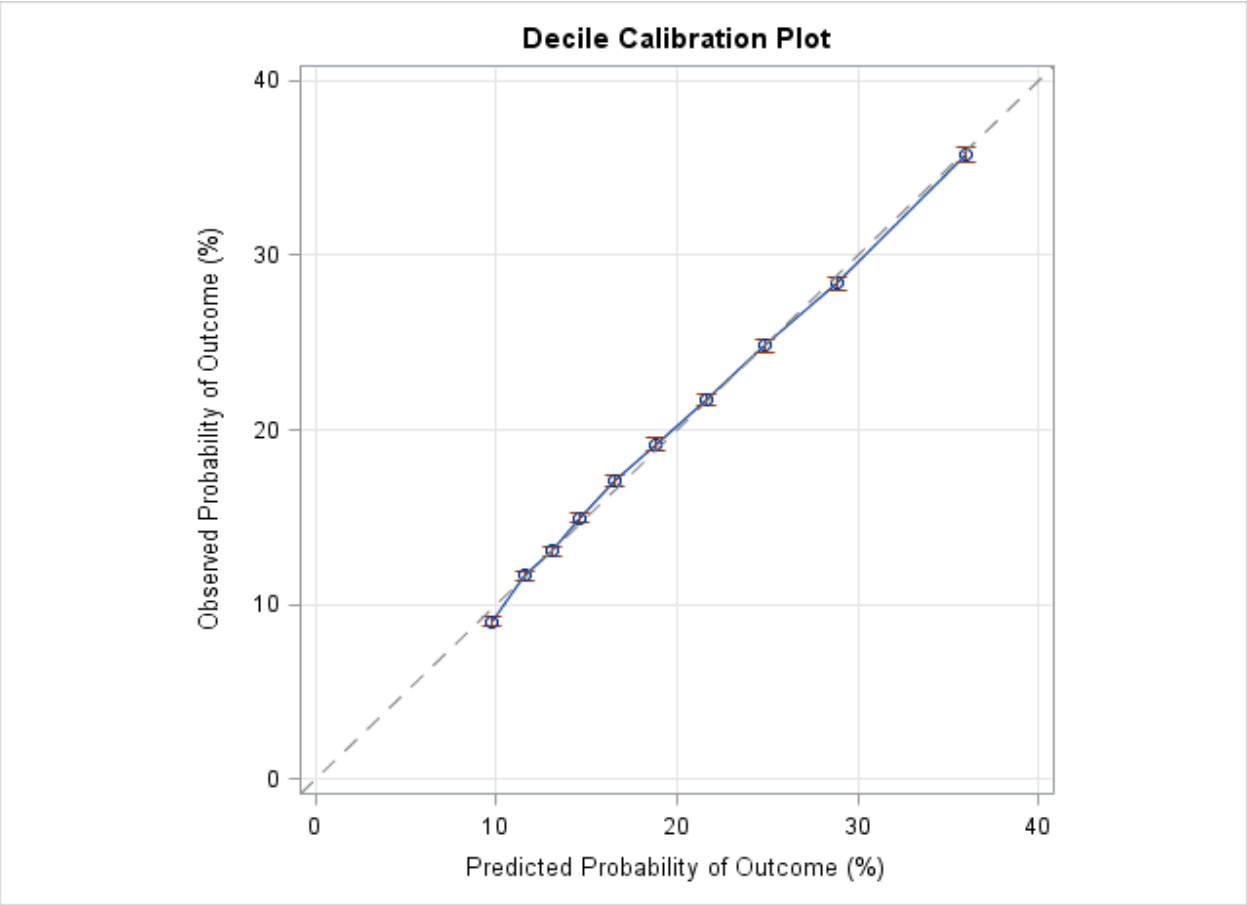
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

The calibration statistics results across the datasets A1, A2 and B indicate the values in close proximity to 0-1, indicating good model reliability. See below for detailed result:

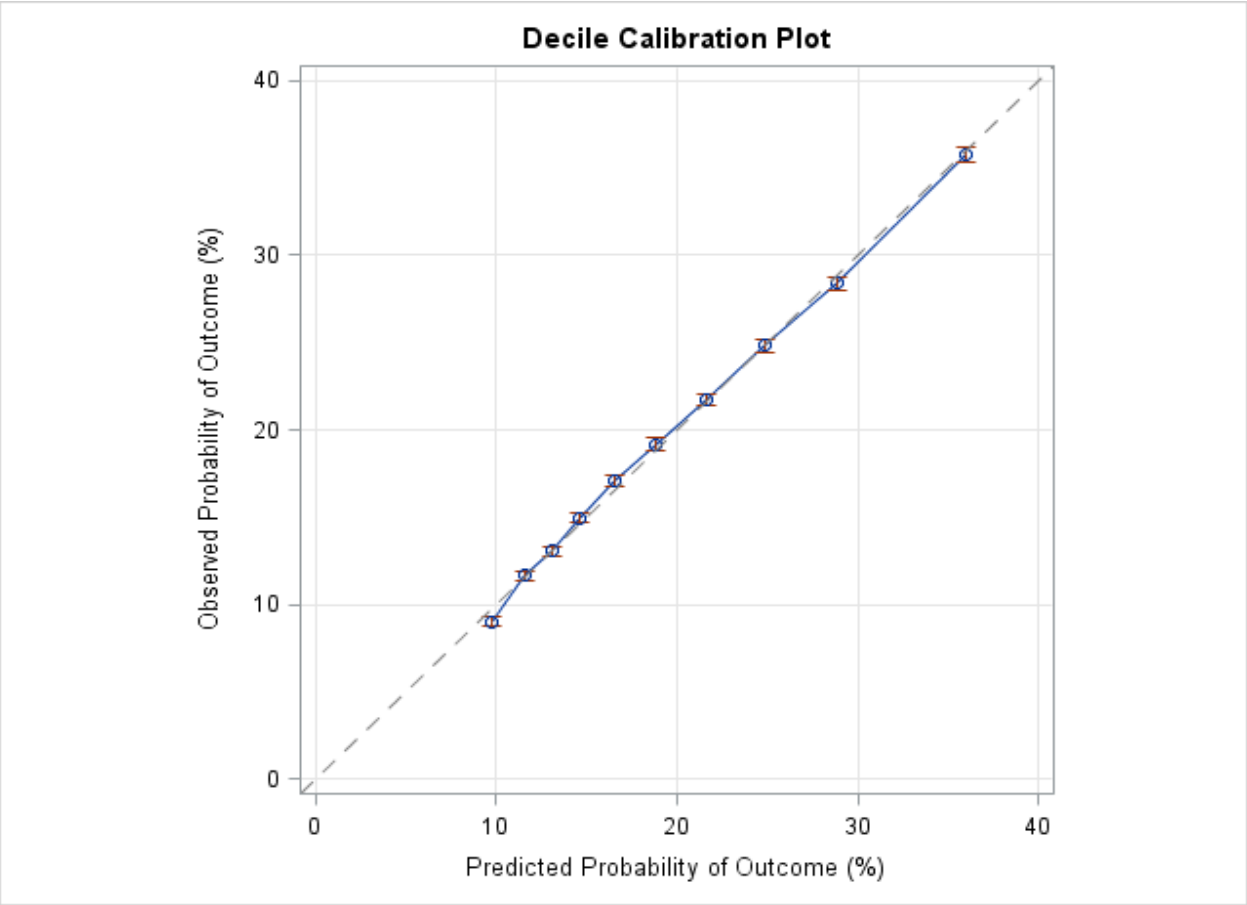
Cohort	2015-2016 Development Sample	2015-2016 Validation Sample	2016-2017 Temporal Validation Sample
Cardiorespiratory	0 - 1	-0.023 - 0.988	-0.023 - 1.002
Cardiovascular	0 – 1	-0.015 - 0.997	-0.018 - 1.001
Medicine	0 – 1	0.000- 1.003	-0.006 - 0.994
Neurology	0 - 1	-0.085 - 0.951	-0.047 - 0.978

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

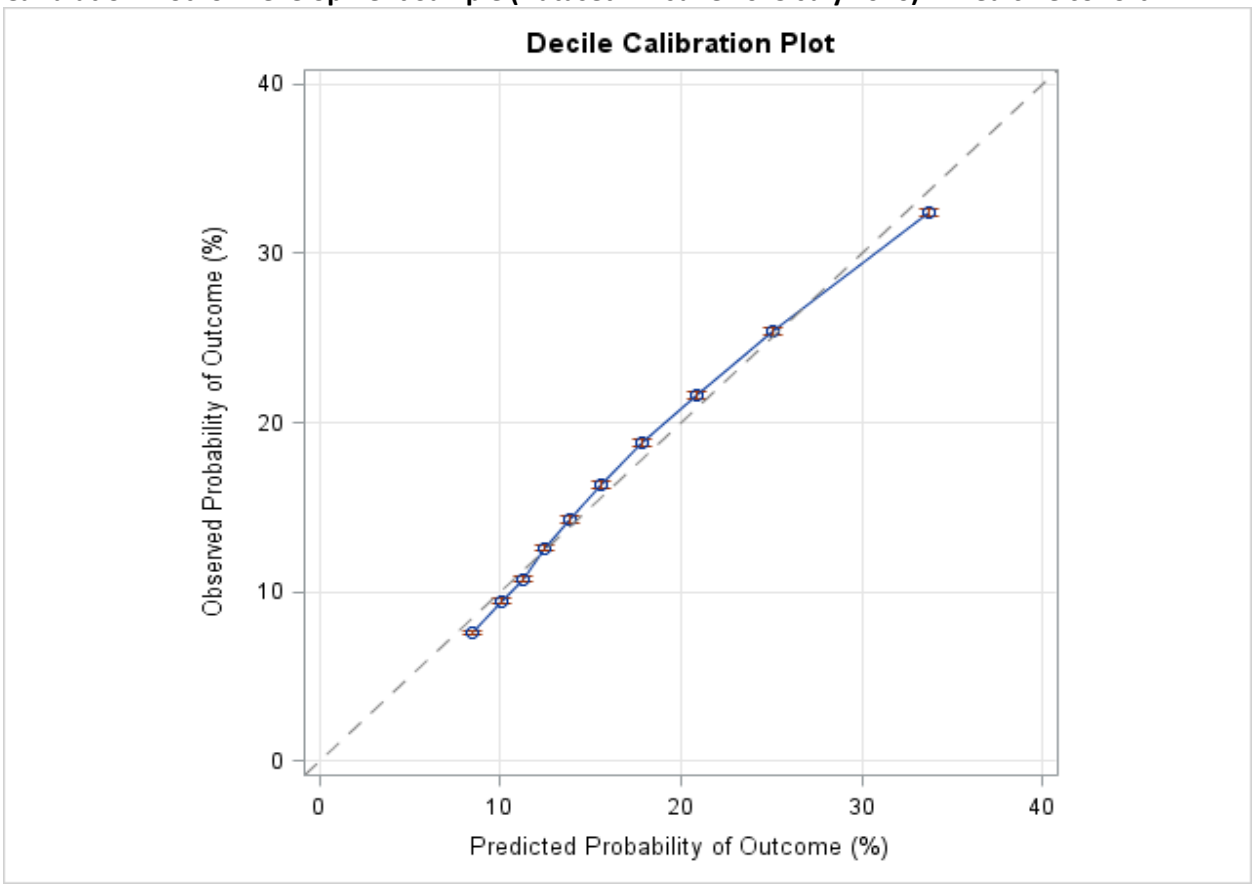
Calibration Plot for Development Sample (Dataset A1: June 2015-July 2016) – CARDIORESPIRATORY cohort



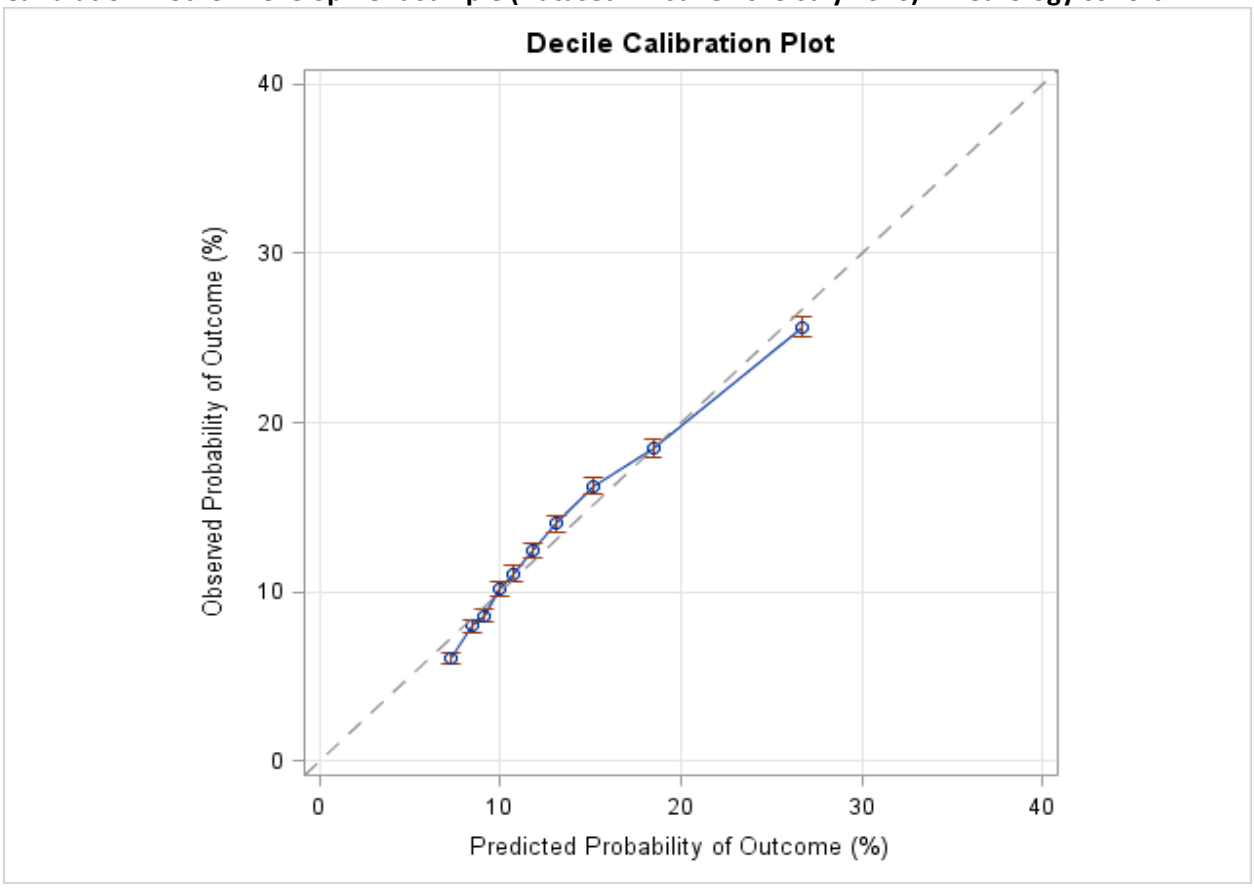
Calibration Plot for Development Sample (Dataset A1: June 2015-July 2016) – CV cohort



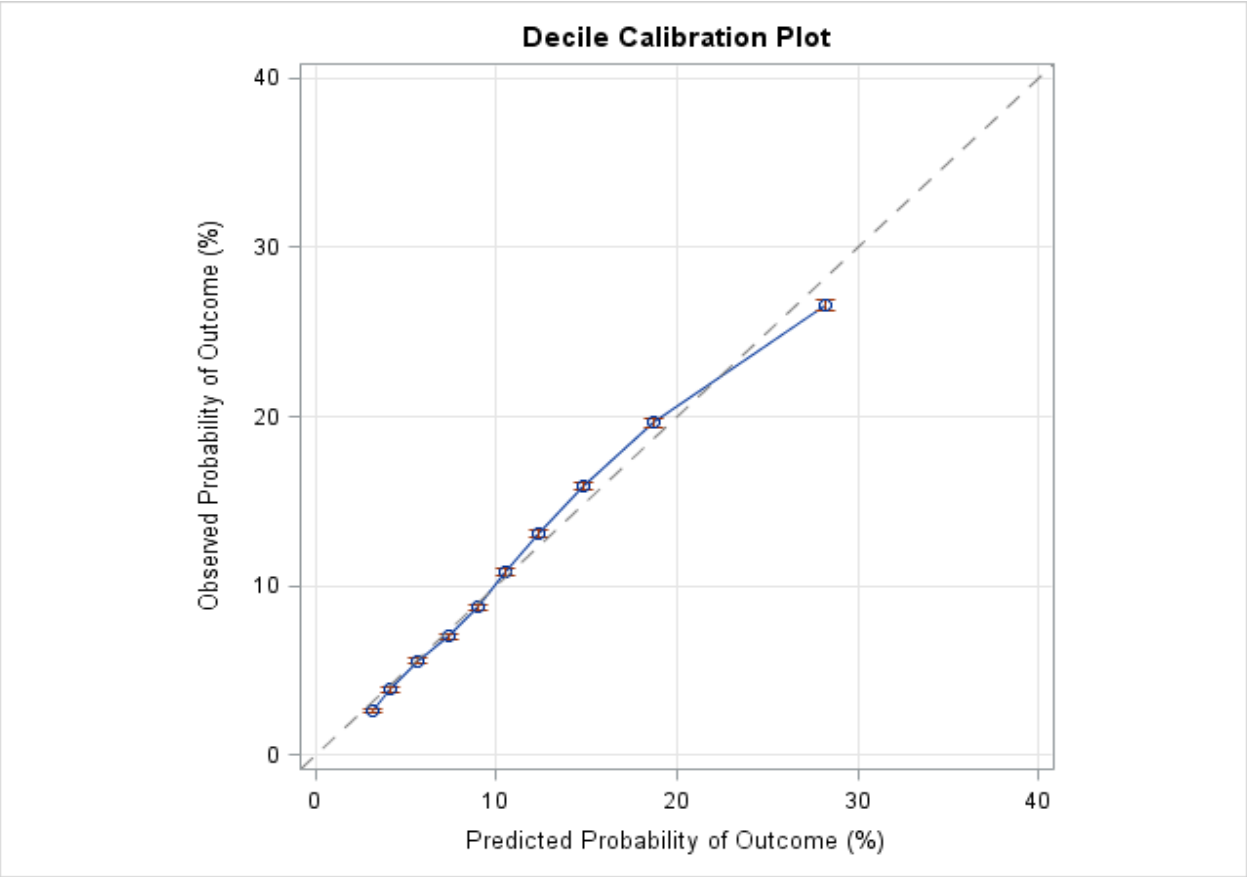
Calibration Plot for Development Sample (Dataset A1: June 2015-July 2016) – Medicine cohort



Calibration Plot for Development Sample (Dataset A1: June 2015-July 2016) – Neurology cohort



Calibration Plot for Development Sample (Dataset A1: June 2015-July 2016) - Surgical cohort



2b3.9. Results of Risk Stratification Analysis:

N/A; the measure is not 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 following results demonstrates that the risk-adjustment model adequately controls for differences in patient characteristics: <u>Discrimination Statistics</u> The calculated c-statistics ranged from 0.63 to 0.71 across specialty cohorts and datasets (Dataset A1, A2 and B); these values indicate good model discrimination across the cohort models. The models also predicted a wide range between the lowest decile and highest decile for each cohort and dataset, indicating the ability to distinguish high-risk subjects from low-risk subjects. <u>Calibration Statistics</u> The calibration values which are consistently of close to 0 at one end and close to 1 for all specialty cohorts and datasets, indicating good calibration of the models across all three datasets. <u>Risk Decile Plots</u> Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates excellent discrimination of the model and good predictive ability.

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

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

N/A

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 provider-level RARR. We characterize the degree of variability by reporting the distribution of the RARR for Eligible Clinician Groups. We also used the bootstrapped 95% confidence intervals to classify Eligible Clinician Groups as being better, worse, or no different than expected, according to whether the interval excluded the overall mean RARR.

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*)

For clinician group level: The risk-adjusted readmission rates estimated using Medicare FFS data (FYs 2015-2016) had a median value of 15.3%. The interquartile values ranged from 14.6% (Q1) to 16.2% (Q3). The percentiles of the distribution were as follows:

Min	1st	5th	10th	25TH	50TH	75TH	90TH	95TH	99TH	Max
7.0%	12.1%	13.3%	13.8%	14.5%	15.3%	16.2%	17.1%	17.7%	19.1%	25.1%

Using a bootstrapped 95% interval estimate, we found 6,447 significant outliers among 55,593 clinician groups. Of the 55,593 clinician groups, 4,318 were categorized as better than expected, 2,129 as worse than expected, and 48,146 as not statistically different than expected.

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

The median RARR is 15.3% which indicates that patients are expected to have readmission in 30 days after discharge on average 15.3% of the time for clinician group level. Further, the 10th and 90th percentiles (13.8 and 17.1) represent meaningful deviations from this median: a clinician group performing at the 10th percentile is performing 9.8% better than an average performer, while a clinician group performing at the 90th percentile is performing nearly 11.8% worse than an average performer. Furthermore, the best performing clinician groups (7.0%) are performing 54.2% better

than an average performer, while the worst performing clinician groups (25.1%) are performing 64.1% worse than an average performer. This variation shows a clear quality gap, as some clinician groups can achieve substantially lower rates than the average performer, while other clinician groups are performing worse than an average performer. It is important to note that here the average performer refers to a clinician group with the same case and procedure mix performing at the average.

We identified a meaningful number of outliers, with more than 10% of eligible clinician groups lying outside the "expected" performance range.

Overall, our results suggest that there is substantial need to both reduce the expected rate and the variation in rates across eligible clinician groups, and that this improvement goal is 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*)

N/A

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

N/A

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

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

Because we used claims data, there were no missing data.

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

N/A

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

N/A

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

N/A

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

Attachment:

3c. Data Collection Strategy

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

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

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

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

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

N/A

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Payment Program	

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

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

N/A

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

N/A

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

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

N/A

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.

N/A

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.

N/A

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

N/A

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

N/A

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.

N/A

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 a primary goal of the measure is to provide information necessary to implement focused quality improvement efforts. Once the measure is implemented, we plan to examine trends in improvements by comparing RSRR 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

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

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

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

This measure is a re-specification of NQF #1789, Hospital Wide All-Cause Unplanned Readmission Measure.

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.

For the NQF #1789 All Cause Unplanned Readmission Measure, attribution is to a facility, with measurement at the hospital level. If used to assess clinician or clinician groups, a facility-based clinician would be assigned the score of the hospital at which the facility-based clinician provides services to the most Medicare patients, and attribution of facility-based groups would be the hospital at which the plurality of facility-based clinicians were attributed. There would be no attribution to outpatient providers. In contrast to facility-based measures, the current measure is an eligible clinician or eligible clinician group-level measure that is aligned with, but not identical to, the original hospital-level measure (#1789). The current measure was developed with input from a diverse Technical Expert Panel that included patients and clinicians to ensure the resulting measure is as meaningful as possible to all stakeholders. The TEP members strongly advocated attributing the measure to multiple clinicians, including outpatient providers, to create incentives for shared accountability for patient readmissions.For the NQF #1789 All Cause Unplanned Readmission Measure, attribution is to a facility, with measurement at the hospital level. If used to assess clinician groups, attribution of facility-based groups would be the hospital at which the plurality of facility-based clinicians were attributed. There would be no attribution to outpatient providers. In contrast to facility-based measures, the current measure is an eligible clinician group-level measure that is aligned with, but not identical to, the original hospital-level measure (#1789). The current measure was developed with input from a diverse Technical Expert Panel (TEP) that included patients and clinicians to ensure the resulting measure is as meaningful as possible to all stakeholders. The TEP

members strongly advocated attributing the measure to multiple clinicians, including outpatient providers, to create incentives for shared accountability for patient readmissions.

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.)

Clinicians, especially those with key roles in caring for the patient, can influence the risk of readmission both directly and through their influence on hospital culture and programs. Therefore, many of the best practices and strategies adopted by hospitals for reducing readmissions can be supported and promoted by clinician groups to improve patient outcomes. Further, by attributing each index admission to multiple clinicians, this measure encourages and incentivizes care coordination among the clinicians with key roles in reducing the risk that the patient returns for unplanned acute care.

Appendix

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

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

Contact Information

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

Co.2 Point of Contact: Lein, Han, Lein.Han@cms.hhs.gov, 410-786-0205-

Co.3 Measure Developer if different from Measure Steward: Yale New Haven Health/Center for Outcomes Research and Evaluation

Co.4 Point of Contact: Lori, Wallace, lori.wallace@yale,edu, 203-200-4478-

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 was comprised of experts in measure development, health services research, clinical medicine, statistics, and measurement methodology. Our measure development team consisted of the following members:

Jeph Herrin, Ph.D. –Project Lead

Lisa G. Suter, M.D. –Project Director

Susannah M. Bernheim, M.D., MHS –Project Director

Faseeha K. Altaf, MPH –Project Coordinator

Ilana B. Richman, M.D. –Clinical Investigator

Elizabeth E. Drye, M.D., S.M. –Clinical Investigator

Shu-Xia Li, Ph.D. –Analyst

Yixin Li, M.S. –Analyst

Zhenqiu Lin, Ph.D. –Analytic Director

Katie Balestracci, Ph.D –Research Scientist

Sriram Ramanan, B.S. –Research Assistant

Rushi Shah, B.S. –Research Assistant

Heather Hussey, MPH –Project Coordinator

Victoria Taiwo, MS–Research Associate

Andreina Jimenez, MPH–Research Associate

Lynette Lines, MS, PMP–Project Manager

Harlan M. Krumholz, M.D., S.M. –Principal Investigator

Technical Expert Panel Members:

Kathleen Blake, MD, MPH

John Birkmeyer, MD

Dale Bratzler, DO, MPH

Daniel Brotman, MD, SFM, FACP

Tracy Cardin, ACNP-BC, SFHM

Cathy Castillo, BA

Bruce Chernof, MD

Donna Cryer, JD

Sherrie H. Kaplan, PhD, MPH

Timothy Kresowik, MD, MS

Joshua Lapps, MA

Frederick Masoudi, MD, MSPH

Brian McCardel, MD

James Moore, MD

Michelle Mourad, MD

Juan Quintana, DNP, MHS, CRNA

Carol Raphael, MA, MPH

Charlene Setlow

Heidi L. Wald, MD, MSPH

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: Please find the full response to S.14. Calculation Algorithm/Measure Logic in the attached Intent to Submit form.

Appendix: Additional information submitted by the developer

Evidence & Attribution

The Hospital-Wide, All-Cause Unplanned Readmission (HWR) Rate for the Merit-Based Incentive Payment Program (MIPS) eligible clinician groups measure is a re-specification of a measure that is currently implemented in the MIPS program. This is the first time that this measure has been considered for endorsement by NQF despite it being used currently in MIPS. The current version in MIPS attributes readmissions solely to the primary care physician that provides the plurality of care during the measurement period, but this may not be a clinician with opportunity to impact readmission (such as clinicians who first see a patient after their readmission).

This new version improves upon the current attribution methodology by considering joint attribution for up to three clinician groups or practices that provide care for patients inside and outside of the hospital prior to discharge and are therefore in positions to influence patients' risk of readmission. This approach to joint attribution was supported by an extensive literature review and stakeholder input from patients, providers, and payers.

This new version of the measure attributes each readmission to:

1. the clinician group of the Primary Inpatient Clinician (PIC),
2. the clinician group of the Discharge Clinician (DC), and
3. the clinician group of the Primary Outpatient Clinician (POC).

Though the same group may be identified as filling two or even three of these attribution roles, the readmission is only attributed once. By holding these three clinical groups jointly accountable, the measure aims to incentivize collaboration of care across inpatient and outpatient settings. Below we list the general responsibilities of each of these clinician groups and how their choices and patterns of care have been shown to influence readmissions. We heard clearly from stakeholders that readmission is a complex clinical outcome requiring multidisciplinary and collaborative care to reduce risk and optimize health outcomes. Clinicians are critical components of quality improvement, whether as change agents and influencers of health systems or key stakeholders whose acceptance is required for any successful or long-lasting improvements, which is why CMS chose to implement readmission measurement into the MIPS program.

The PIC group is defined as the clinician group for the PIC for a given patient - that is, the clinician who billed the most charges for the patient during their hospital say. Such clinicians are most likely responsible for ensuring relevant medical problems are addressed in the inpatient setting, reducing the chance patients will return to the hospital with unresolved medical issues. For example, a medical specialty PIC is the clinician likely seeing a patient on a regular basis while a surgical specialty PIC likely performed a significant surgery for the patient. These clinicians, and the clinician groups they are part of, represent one of the three clinician groups to which this measure attributes readmissions. They are likely to make decisions about what medications the patient needs and what other specialties or providers should be involved in the patient's hospital care. Studies have shown that selection of affordable medications with favorable side effect profiles and dosing schedules that are easiest to adhere to, such as daily or twice daily dosing, predict adherence and lower readmission.

The Discharge Clinician (DC) group is defined as the clinician group for the DC for a given patient - that is, the clinician transitioning the patient from inpatient to outpatient care. The Discharge Clinician (DC) is most responsible for preparing a patient for discharge. Responsibilities for the DC include determining that the patient is, in fact, clinically appropriate to leave the hospital and ensuring they understand their medical condition and treatments. The DC can also directly support and influence other providers and staff to support critical actions related to discharge: providing hard copies of discharge instructions in a language the patient can understand, ensuring the instructions are concordant with what was communicated to the patient by the PIC and other clinicians during the inpatient stay, and verbal explanation of the discharge instructions including management of medications, referrals to outpatient specialists or therapy, and lifestyle

modifications, all of which ensure the patient adheres to the plan of care upon discharge from the hospital.[1,2] Studies have shown that interventions focused on inpatient clinicians such as the PIC and DC can strengthen discharge systems by making phone calls, scheduling appointments, and providing information on the transition to a Primary Care Provider – these have all been shown to improve patient access to a clinician once they are discharged from the hospital and reduce readmission. [3,4]

The Primary Outpatient Clinician (POC) group is defined as the clinician group for the POC for a given patient - that is, the clinician with the greatest number of claims for primary care during the 12 months prior to the hospital admission rate. This clinician can reduce the chance that a patient will be readmitted by having open access and ensuring available appointments for the patient within 30 days of discharge.

Attribution to these different providers was vetted by the Technical Expert Panel (TEP), which was comprised of hospitalists and quality measure experts, during the development of the MIPS HWR measure. The TEP considered the fact that readmission is a multi-factorial outcome and that multiple clinicians may play important roles in providing appropriate care, making good decisions and practical recommendations, and intentionally planning care transitions. These actions, while only part of the overall success of a care transition, are necessary to promote the best possible outcome. In this context, most TEP members agreed that patient outcomes should not be attributed to a single clinician, but should rather be attributed to multiple clinicians responsible for the patient during and after the inpatient stay. Some TEP members noted the multiple attribution algorithm addressed the current reality of patient care, as opposed to being optimistic in how care should be coordinated. Some TEP members also felt that clinicians in the hospital were best able to drive system changes. TEP members were also supportive of attributing to the POC as they felt it was in part the responsibility of the outpatient provider to ensure the patient does not return to the hospital with remnant issues from their inpatient stay. To be sure we heard the voice of the outpatient provider, we also reviewed the attribution approach with CMS' QPP Clinical Champions through a call with outpatient primary care providers and providers who provide primary care in both the inpatient and outpatient settings.

While it is possible for a clinician group to include a single clinician, it is also worth noting that this measure is intended for reporting within the Merit-Based Incentive Payment System (MIPS) program, which only includes clinician groups, or TINs, with at least 16 clinicians, or NPIs. Thus, even if the same individual clinician served in all three attributed roles (PIC, DC, and POC), the outcome would still only be attributed to their group if it includes at least 16 different clinicians, and not to that individual.

SDS Adjustment

Literature referenced in the testing attachment indicates that the relationships between patient social risk factors and readmissions are multifaceted. Causal pathways include patient health upon admission, social risk factors outside of the hospital, care quality of the hospital, and differential care within a hospital. While this measure attributes readmissions to clinician groups, it still captures quality care provided by clinicians within the hospital system. Thus, the conceptual framework used to understand how social risk factors play a role in

---

[1] Bowles KH, Hanlon A, Holland D, Potashnik SL, Topaz M. Impact of discharge planning decision support on time to readmission among older adult medical patients. *Prof Case Manag*. 2014;19(1):29–38. doi:10.1097/01.PCAMA.0000438971.79801.7a

[2] Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis [published correction appears in JAMA. 2004 Sep 1;292(9):1022]. *JAMA*. 2004;291(11):1358–1367. doi:10.1001/jama.291.11.1358

[3] DeCaporale-Ryan LN, Ahmed-Sarwar N, Upham R, Mahler K, Lashway K. Reducing hospital readmission through team-based primary care: A 7-week pilot study integrating behavioral health and pharmacy. *Fam Syst Health*. 2017;35(2):217–226. doi:10.1037/fsh0000269

[4] Verhaegh KJ, MacNeil-Vroomen JL, Eslami S, Geerlings SE, de Rooij SE, Buurman BM. Transitional care interventions prevent hospital readmissions for adults with chronic illnesses. *Health Aff (Millwood)*. 2014;33(9):1531–1539. doi:10.1377/hlthaff.2014.0160

how hospitals influence readmissions can also be used to understand how social risk factors play a role in how clinician groups at the hospital can influence readmissions.

Of the pathways listed, overall quality of hospitals and differential care within a hospital are captured by readmission outcome measures, while health status upon admission is accounted for through risk adjustment for patient case mix and hospital service mix. To address the potential impact of social risk factors outside of the hospital, we tested for the effects of including two social risk factors within the model (dual eligibility status and low Agency for Healthcare Research and Quality SES) on final risk-adjusted rates for clinician groups. We considered these because they are available and reliably measured for all Medicare beneficiaries in the cohort. Ongoing research aims to identify valid patient-level social risk factors and highlight disparities related to social risk, and as additional variables become available, they will be considered for testing and inclusion within the measure. We chose not to include either of these social risk factors for two reasons.

First, for both social risk factors, the correlation between the adjusted and unadjusted scores was 0.99, indicating extremely high agreement. This supports that adding these social risk factors would have minimal impact on measure scores. In addition to the correlation between adjusted and unadjusted scores, we also tested the absolute change in risk-adjusted readmission rates. When incorporating the dual eligible risk factor, the largest change in a hospital-level risk-adjusted readmission rate was 1.74% for clinician groups. When incorporating low AHRQ SES, the largest change in a hospital-level risk-adjusted readmission rate was 2.45%.

For risk-adjusted outcome measures, CMS' standard approach is to first consider adjustment for clinical conditions and then examine additional risk imparted by social risk factors after the potential for greater disease burden is included in the risk model. We believe that this is consistent with NQF current guidance and is appropriate given the evidence cited in our submission that people who experience greater social risk are more likely to have more disease burden compared with those who do not. In addition, according to NQF guidance, developers should assess social risk factors for their contribution of unique variation in the outcome – that they are not redundant.[5] Therefore, if clinical risk factors explain all or most of the patient variation in the outcome, then NQF guidance does not support adding social risk factors that do not account for variation. Based on the results reported in the prior paragraph, this approach and guidance suggests that there is not a strong justification for including the social risk factors.

Second, one of the key principles behind development of this measure was to align when possible with the original hospital Hospital-Wide Readmission measure reported within IQR. As that measure (NQF #1789) doesn't adjust for social risk factors, we would suggest that the precedent argues for omitting them here as well.

[3] National Quality Forum (NQF). Risk adjustment for socioeconomic status or other sociodemographic factors: Technical report. 2014; http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx. Accessed September 3, 2019.